IJCSIS Vol. 14 No. 6, June 2016 Part 1
ISSN 1947-5500

International Journal of Computer Science & Information Security

Indexed and technically co-sponsored by :

IJCSIS

ISSN (online): 1947-5500

Please consider to contribute to and/or forward to the appropriate groups the following opportunity to submit and publish original scientific results.

CALL FOR PAPERS
International Journal of Computer Science and Information Security (IJCSIS)
January-December 2016 Issues

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.
See authors guide for manuscript preparation and submission guidelines.

Indexed by Google Scholar, DBLP, CiteSeerX, Directory for Open Access Journal (DOAJ), Bielefeld Academic Search Engine (BASE), SCIRUS, Scopus Database, Cornell University Library, ScientificCommons, ProQuest, EBSCO and more.

Deadline: **see web site**
Notification: see web site
Revision: see web site
Publication: see web site

Context-aware systems
Networking technologies
Security in network, systems, and applications
Evolutionary computation
Industrial systems
Evolutionary computation
Autonomic and autonomous systems
Bio-technologies
Knowledge data systems
Mobile and distance education
Intelligent techniques, logics and systems
Knowledge processing
Information technologies
Internet and web technologies
Digital information processing
Cognitive science and knowledge
Agent-based systems
Mobility and multimedia systems
Systems performance
Networking and telecommunications
Software development and deployment
Knowledge virtualization
Systems and networks on the chip
Knowledge for global defense
Information Systems [IS]
IPv6 Today - Technology and deployment
Modeling
Software Engineering
Optimization
Complexity
Natural Language Processing
Speech Synthesis
Data Mining

For more topics, please see web site https://sites.google.com/site/ijcsis/

For more information, please visit the journal website (https://sites.google.com/site/ijcsis/)

Editorial
Message from Editorial Board

*It is our great pleasure to present the **June 2016 issue** (Volume 14 Number 6 Part 1, 2 & 3) of the **International Journal of Computer Science and Information Security** (IJCSIS). High quality research, survey & review articles are proposed from experts in the field, promoting insight and understanding of the state of the art, and trends in computer science and technology. It especially provides a platform for high-caliber academics, practitioners and PhD/Doctoral graduates to publish completed work and latest research outcomes. According to Google Scholar, up to now papers published in IJCSIS have been cited over 6390 times and the number is quickly increasing. This statistics shows that IJCSIS has established the first step to be an international and prestigious journal in the field of Computer Science and Information Security. There have been many improvements to the processing of papers; we have also witnessed a significant growth in interest through a higher number of submissions as well as through the breadth and quality of those submissions. IJCSIS is indexed in major academic/scientific databases and important repositories, such as: Google Scholar, Thomson Reuters, ArXiv, CiteSeerX, Cornell's University Library, Ei Compendex, ISI Scopus, DBLP, DOAJ, ProQuest, ResearchGate, Academia.edu and EBSCO among others.*

On behalf of IJCSIS community and the sponsors, we congratulate the authors and thank the reviewers for their outstanding efforts to review and recommend high quality papers for publication. In particular, we would like to thank the international academia and researchers for continued support by citing papers published in IJCSIS. Without their sustained and unselfish commitments, IJCSIS would not have achieved its current premier status.

*"We support researchers to succeed by providing high visibility & impact value, prestige and excellence in research publication." For further questions or other suggestions please do not hesitate to contact us at **ijcsiseditor@gmail.com**.*

A complete list of journals can be found at:
http://sites.google.com/site/ijcsis/

IJCSIS Vol. 14, No. 6, June 2016 Edition

ISSN 1947-5500 © IJCSIS, USA.

Journal Indexed by (among others):

Bibliographic Information
ISSN: 1947-5500
Monthly publication (Regular Special Issues)
Commenced Publication since May 2009

Editorial / Paper Submissions:
IJCSIS Managing Editor
(ijcsiseditor@gmail.com)
Pennsylvania, USA
Tel: +1 412 390 5159

IJCSIS EDITORIAL BOARD

TABLE OF CONTENTS

1. PaperID 31051608: ESSPI: Exponential Smoothing Seasonal Planting Index, a New Algorithm for Prediction Rainfall (pp. 1-9)

Kristoko D. Hartomo, Faculty of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia
Subanar, Faculty of Mathematics and Natural Sciences, GadjahMada University, Yogyakarta, Indonesia
Edi Winarko, Faculty of Mathematics and Natural Sciences, GadjahMada University, Yogyakarta, Indonesia

Abstract — Exponential smoothing algorithm is a prediction algorithm recommended by the Food and Agriculture Organization. The weakness of exponential smoothing prediction algorithm is low accuracy for the prediction of long-term and ineffective in determining the value of smoothing to minimize error. The proposed research is to build a model rainfall prediction using a new algorithm Seasonal Planting Index (ESSPI). By using the algorithm planting seasonal index, rainfall prediction model will generate higher accuracy. The results showed seasonal planting method is the dominant index (5 of 6 test size) have an average accuracy is better than the method of exponential smoothing. Index planting seasonal prediction accuracy of 95.73% better than the exponential smoothing $\alpha = 0.1$ by 56.55%, and exponential smoothing of $\alpha = 55.53$. Novelty of this research is new algorithms for classifying data based on seasonal planting index, a new algorithm for determining the smoothing (value), the new fitting algorithm using seasonal planting index, and new algorithms using seasonal rainfall prediction planting index for the determination of the growing season.

Keywords—exponential; smoothing; algorithm; seasonal planting index; predictions; accuracy; rainfall; novelty

2. PaperID 31051609: A New MultiPathTCP Flooding Attacks Mitigation Technique (pp. 10-15)

Adwan Yasin, Department of Computer Science, Arab American University, Jenin, Palestine
Hamzah Hijawi, Department of Computer Science, Arab American University, Jenin, Palestine

Abstract — MPTCP is a new protocol proposed by IETF working group as an extension for standard TCP, it adds the capability to split the TCP connection across multiple paths. It provides higher availability and improves the throughput between two multi-address endpoints. Many Linux distributions have been developed to support MPTCP, most of them are open source which can be modified and compiled to support different experimental scenarios. Splitting the single path TCP connection across multiple paths adds new challenges in paths management and raises new security threats. Some of these threats include flooding and hijacking attacks performed by on-path and offpath attackers. In this article, we propose a new algorithm to mitigate the flooding and hijacking attacks in MPTCP, the proposed method allows a stateful processing of the initial SYN message and it's following SYN_JOIN messages.

Keywords — TCP, MPTCP, flooding, hijack, on-path, off-path, flooding, DoS

3. PaperID 31051613: Temporal Performances Evaluation of Multi-Robot Demining System Inspired by Ant Behavior (pp. 16-24)

Riadh SAAIDIA, Mohamed Sahbi BELLAMINE, Abdessattar BEN AMOR
Computer Laboratory for Industrial Systems (LISI), National Institute of Applied Sciences and Technology (University of Carthage), INSAT, TUNISIA

Abstract — In this paper we adopt a cooperative strategy based on ACO (Ant Colony Optimization) algorithms to coordinate a Multi Robots System (MRS). Our principal objective is to evaluate temporal performances for this system by choosing demining operations as a benchmark problem. In this work, we try to adapt the ACO algorithm parameters for different mine distribution in order to reduce time demining operations. In particular, we report effects of evaporation pheromone rate model and minefield configuration on temporal performances.

Index Terms— ACO algorithms, multi-robot system (MRS), evaporation pheromone rate, demining system.

4. PaperID 31051614: Towards Developing a Cost Effective Solution for Environmental Monitoring (pp. 25-28)

Muhammad Soban Khan, Ans Ali Raza, Zeeshan Musawar, Shoaib Hassan, Taimoor Hassan
Department of Computer Science, COMSATS Institute of Information Technology, Sahiwal, COMSATS Road off GT road, Sahiwal 57000, Pakistan

Abstract - Environment refers to everything that surrounds a person. Environment contains many types of pollution. Most dangerous pollution is air pollution. Most important factor that causes human health is air pollution. Many countries are suffering from air pollution. There are many factors that cause air pollution. Some major factors are smoke, carbon monoxide and high temperature. Many developing countries are creating solutions for detecting and analyzing the air pollution. The main idea of our research is based on proposing a cost effective solution for environmental detection. Our system is a connection between sensors, Raspberry Pi, Microsoft Azure and Android Mobiles. Raspberry Pi gets environmental values with help of Raspberry Pi and sends the data to Microsoft Azure through API, form where Android Mobile gets those values with the help of HTTP request. Our proposed system successfully detected temperature, humidity, hydrogen, methane, propane, carbon monoxide and air level. The results show that our system is most cost effective, secure and easy to use. It will helpful in saving lives.

Keywords: Environment Pollution, Environmental monitoring system, Raspberry Pi, Air pollution

5. PaperID 31051615: AV Encryption Algorithm to Protect Audio visual Content for IPTV (pp. 29-39)

Muhammad Akram, C. A. Rahim, Amjad Hussain Zahid
The Institute of Management Sciences (PAK-AIMS), 54660 Lahore, Pakistan

Abstract — Crypt analytical techniques for multimedia technologies particularly audio visual applications have shown some existing flaws while maintaining the security and computational time. This case study is a representative algorithm especially for protection of IPTV contents. The network's reliability and security of contents is the major issue in IPTV media business. The proposed algorithm is the Audio Video MPEG file encryption technique in which the synchronization between audio and video and the frame sequence is shuffled before the transmitting end or vertical device. . The shuffling process is guided by input key frames to point out frame positions. The MPEG video frames are first extracted via spatial pyramid kernel. It divides the stream into regions over different scales and to find out the frame similarity while on merging of AV frames. Then ciphers are implemented to locate the shuffled frames and further genetic algorithm such as AES is used to encrypt. By this way, AV contents of IPTV can be secure from malicious users.

Keywords— MPEG, IPTV, CAS, DRM, DES, AES

6. PaperID 31051616: Secure Speaker Biometric System using GFCC with Additive White Gaussian Noise and Wavelet Filter (pp. 40-47)

Gaganpreet Kaur, Deptt. of CSE, I.K. Punjab Technical University, Punjab, India
Dr. Dheerendra Singh, Deptt. of CSE, Chandigarh College of Engineering and Technology, Sector-26, Chandigarh, India

Abstract — Speaker Identification (SI) aims to identify the speaker’s identity from the given list of speakers. Speaker identification is efficient under the clean training and testing environment conditions. In real environment application, there occurs mismatch between training and testing environments due to background noise, which degrades the system’s performance and security. So, robust speaker identification is the important issue in research. This paper

describes the recently used front end algorithm based on Gammatone Frequency Cepstral Coefficients (GFCC) along with speech detection algorithm and Cepstral mean normalization (CMN). System makes model using Gaussian Mixture Model (GMM) Classifier, which uses iterative Expectation Maximization (EM) Algorithm to estimate the Gaussian model parameters. Training data is taken in clean environment and all test utterances are corrupted by adding White Gaussian Noise (AWGN). This paper aims to improve the robustness of speaker identification even when additive noise is added during testing phase. For improvement Wavelet Filter is implemented to de-noise the speech signal. Experiment is carried out in real database oriented and stored database oriented relative to the Attendance System application. Experiment is carried on 100 speakers saying phrases like 'Yes mam' "present mam", 'Yes sir', 'present sir' with 4 types of utterances for each phrase (so database includes 400 utterances). Experiment results obtained shows better performance in noisy environment. The results for stored database oriented experiment show that the algorithm gives 85% of Correct Recognition Rate (CORR) while using wavelet filter and 73% without using the filter. The results for real database oriented experiment shows 74% of identification rate while using wavelet filter and 45% without using the filter.

Keywords — Gammatone Frequency Cepstral Coefficients (GFCC); Gaussian Mixture Model (GMM); Cepstral mean normalization (CMN); Robust Speaker Identification, Additive White Gaussian Noise (AWGN); Wavelet Filter.

7. PaperID 31051620: A Novel Algorithm for Load Balancing using HBA and ACO in Cloud Computing Environment (pp. 48-52)

Seyed Majid Mousavi, University of Debrecen, Faculty of Informatics, Debrecen, Hungary
Fazekas Gábor, University of Debrecen, Faculty of Informatics, Debrecen, Hungary

Abstract — Cloud computing is an emerging technology and new trend for computing based on virtualization of resources. Scheduling of tasks to reach load balancing is a challenge in cloud environment. Load balancing is the process of distribution of the load among VMs in order to efficiently utilize of resources and avoiding the situation where some VMs are overloaded or idle. Load balancing of non-preemptive tasks is one of the critical issues in task scheduling in clouds environment. To improve throughput at cloud resources, an intelligent and dynamic load balancing can significantly increase cloud's performance and minimize the costs. Although, many algorithms, strategies and methods have been proposed, but load balancing is still one of the challenging issues in resource allocation in cloud computing environment. In this paper we propose a novel load balancing strategy using Honey Bees and Ant Colony behavior algorithms in cloud environment. The proposed algorithm strives to balance the load of the virtual machines, trying to minimize the completion time of given tasks and reduce response time in cloud infrastructure.

Keywords: load balancing, ant colony, honey bee, cloud computing.

8. PaperID 31051621: Route Optimization in MANET Using Hopfield Neural Networks: MANET-HOP (pp. 53-59)

Sanjeev Gangwar, Department of Computer Application, V. B. S. Purvanchal University, Jaunpur, India
Dr. Krishan Kumar, Department of Computer Science, Gurukul Kangri University, Haridwar, India

Abstract — As we know that Mobile Ad Hoc Network is the combination of nodes having unstable setup which usually formed instantly in independent manner. It does not have any centralized administration. Moreover they don't have any permanent setup and routers. In such situations routing becomes the responsibility of individual nodes and also routing is equally important to realize the practical benefits of MANET. Traditional protocols of MANET: DSR, AODV, DSDV, OLTP work well but still need improvements time-to-time as per the new issues like QoS provisioning and routing. Above protocols mainly depends on hop count measurement. In this paper we have implemented a specific problem of six nodes situated at different locations with primary goal to find the shortest route visiting each node at least once which is based on the concept of Travelling Salesman Problem using Feedback/Hopfield Neural Network. And we found that Hopfield networks are suitable to find the shortest route.

Keywords- Mobile ad-hoc network, Hopfield neural network, Travelling salesman problem, Route optimization

9. PaperID 31051629: A Modified Black hole-Based Task Scheduling Technique for Cloud Computing Environment (pp. 60-67)

Fatemeh Ebadifard, Department of computer, Iran University of science and technology, Tehran, Iran
Zeinab Borhanifard, Department of computer,Qom University, Qom, Iran
Ahmad Akbari, Department of computer, Iran University of science and technology, Tehran, Iran

Abstract — The issue of scheduling is one of the most important ones to be considered by providers of the cloud computing in the data center. Using a suitable solution lets the providers of cloud computing use the available resources more. Additionally, the satisfaction of clients is met through provision of service quality parameters. Most of the solutions for this problem aim at one of the service quality factors and in order to achieve this goal, variety of methods are used. Using the algorithm of modified black hole in this paper, a proper solution is presented to tackle the problem of scheduling the affairs in cloud environment. The proposed method reduces makespan, increases degree of load balancing, and improves the resource`s utilization by considering the capability of each virtual machine. We have compared the proposed algorithm with existing task scheduling algorithms. Simulation results indicate that the proposed algorithm makes a good improvement regarding the makespan and amount of resource utilization compared to schedulers based on Random assignment and particle swarm optimization Algorithms.

Keywords- cloud computing; task scheduling; Black hole; makespan; resource utilization.

10. PaperID 31051631: A Multicast Routing Protocol Based on ODMRP with Stable link in Mobile Ad Hoc Networks (pp. 68-75)

Ebrahim Asadi, Department of Computer Engineering, Shabestar Branch,Islamic Azad University, Shabestar, Iran
Ali Ghaffari, Department of Computer Engineering, Shabestar Branch,Islamic Azad University, Shabestar, Iran

Abstract — Mobile ad hoc networks are more flexible than tradition networks since they do not require fixed infrastructure and allow all nodes move in a random trajectory, which leads frequent rerouting and degrades network performance. So, an important issue in mobile computer network research is routing in mobile ad hoc networks. Multicast sending is one of the methods used for routing in mobile ad hoc networks because of its group activities. However, some problems exist in multicast sending. For example, when receiver nodes attempt to send acknowledgments or path repetition packets simultaneously, crashes may occur, which leads to packet loss. On the other hand, link expiration is another reason for packet loss. In this study, a multicast routing protocol is offered, which uses a combination of two parameters of the received signal's power and the remaining energy to estimate the stability of the link. SINR is used at each node in conjunction with various transmitters to determine a reliable path that reduces link failure and end-to-end delay. The aim is to find the best link with probability of the highest life cycle for each path. Simulation results of the proposed method using NS-2 simulator indicate the good performance of IMP-ODMRP measures in packet delivery rate, end-to-end delay, packet loss rate, and packet collision rate.
Keywords-Mobile ad hoc networks; multicast; routing; IMP-ODMRP protocol; Standard ODMRP; Stable Link.

11. PaperID 31051639: A Survey on Human Social Phenomena inspired Algorithms (pp. 76-81)

Thanh Tung Khuat, My Hanh Le
DATIC Laboratory, IT Faculty, University of Science and Technology – The University of Danang, Vietnam

Abstract — The problem of seeking the optimal solution in the field of science and engineering has been becoming complex and challenging due to the explosion of dimensions and the interdependence of variables. Over the past few decades, a variety of new concepts, techniques and computational applications inspired from nature have been proposed and used to deal with a wide range of optimization problems in diverse fields. Many of nature-inspired algorithms generate high-quality solutions for real-world optimization tasks. Nevertheless, the majority of these

methods are inspired by either biological phenomena or social behaviors of mainly animals and insects. There are few works relied on social phenomena of human being used to form optimization algorithms. This paper aims at presenting an adequate review of most predominant and successful groups of optimization approaches based on human social phenomena.

Index Terms—Human Social Phenomena, Society Civilization Algorithm, Cultural Algorithms. Teaching-learning-based Optimization, Social Learning Algorithm, Alliance Formation based Algorithms, Social Emotional Optimization Algorithm, Social Labeling.

12. PaperID 31051641: Mammogram Classification Using Selected GLCM Features and Random Forest Classifier (pp. 82-87)

Vibhav Prakash Singh, Ayush Srivastava, Devang Kulshreshtha, Arpit Chaudhary, Rajeev Srivastava
Department of Computer Science & Engineering, Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh-221005, India

Abstract - Early diagnosis of breast cancer can improve the survival rate by detecting the cancer at initial stage. Mammogram is a low dose X-ray image of the breast region, used to diagnose the breast cancer at early stage. In this paper, an efficient computer added diagnosis (CAD) system is proposed, automatically detects the normal and abnormal images of mammogram. The proposed pre-processing steps include, cropping of mammograms (for avoiding the pectoral muscle, unwanted tags) and suppression of Gaussian noise. Further, gray level co-occurrence matrix (GLCM) based statistical texture feature from different distances of neighboring and angles are extracted. Furthermore, most relevant features are also examined using AdaBoost feature selection method. Finally, normal and abnormal mammograms are classified using Random forest (RF) classifier. Experiments on benchmark mammography image analysis society (MIAS) database confirm the effectiveness of this work.

Keywords-CAD; Mammography; GLCM features; Feature selection; Random forest classifier.

13. PaperID 31051643: Enhancement of Intrusion-Detection System in MANETs with the Digital Signature via Elliptic Curve Cryptosystem (pp. 88-94)

K. Spurthi, T. N. Shankar, S. Sabari Giri Murugan
Computer Science & Engineering, KL University, AP, India

Abstract- The watchdog scheme is popular in MANET to defend the malicious attacks, but the major pitfall of this method is unable to detect some destructive actions. The technique Enhanced adaptive acknowledgment EAACK is designed to handle some weaknesses as false misbehavior, limited transmission power, and receiver collision of the watchdog scheme that is not fully efficient to resolve all the problems. This paper focuses intrusion detection system on MANETs with the collaboration of three IDS approach and with the techniques ACK, 2-ACK, and misbehavior report identification MRI. This paper proposes digital signature with Elliptic Curve Cryptosystem to avoid forging acknowledgment packets from attackers.

Keywords: DSR, MANET, AOMDV, watchdog, ACK, 2-ACK, MRI.

14. PaperID 31051644: P-Method: Improving AODV Routing Protocol for Against Network Layer Attacks in Mobile Ad-Hoc Networks (pp. 95-103)

Shahram Zandiyan, Department of Computer Engineering, Ardabil branch, Islamic Azad University, Ardabil, Iran
Reza Fotohi, Department of Computer Engineering, Germi branch, Islamic Azad University, Germi, Iran
Marzieh Koravand, Department of Computer Engineering, Germi branch, Islamic Azad University, Germi, Iran

Abstract — Mobile ad hoc networks are regarded as a group of networks consisted of wireless systems which developing together a network with self-arrangement capability. no constant communication infrastructure and use central nodes to communicate with other nodes. Despite lots of advantages, these networks face severe security challenges, since their channels are wireless and each node is connected to central node. One of these concerns is the incidence of network layer attacks (Black and worm hole attack) is one kind of routing disturbing attacks and can bring great damage to the network. In this attack, an attacker cheats nodes, absorbs their packets and then deletes them. Hence, black hole and wormhole disrupts communication, or even makes it impossible in some cases. In this paper, we proposed P-Method for against network layer attacks in mobile Ad-Hoc networks based on hop count and RTT test. The proposed algorithm is implemented in ns2.35 environments and is compared with AODV And DSR under attacks, and improved AODV in different scenarios. Simulation results revealed that the (P-method), is better than AODV And DSR under attack in terms of packet dropped, packet loss, throughput, and jitter.

Keywords- Mobile ad hoc networks, AODV and DSR routing protocol, Black hole attack, Worm hole, P-Method.

15. PaperID 31051653: Check the Use of Raise in Wireless Sensor Networks Based on Heuristic Algorithms Along with Soft Computing Approach (pp. 104-119)

Abolfazl Akbari, Department of Computer Engineering, Ayatollah Amoli Branch, Islamic Azad University, Amol, Iran
Pourya Khodabandeh, Marlik Higher Education Institute, Nowshahr, Iran
Ali Khosrozadeh, Department of Computer Engineering, Ayatollah Amoli Branch, Islamic Azad University, Amol, Iran

Abstract - The use of Wireless Sensor Networks (WSNs) has grown dramatically in recent decades, and the use of these networks in the areas of military, health, environment, business, etc. increases every day. A wireless sensor network consists of many tiny sensor nodes with wireless communications and work independently. In applications of such sensor nodes, hundreds or even thousands of low-cost sensor nodes are dispersed over the monitoring area, in which each sensor node periodically reports its sensed data to the base station (sink). Due to limitations in the communication range, sensor nodes transmit their sensed data through multiple hops. Each sensor node acts as a routing element for other nodes for transmitting data. One of the most important challenges in designing such networks is the management of energy consumption of nodes; because replacing or charging the batteries of these nodes are usually impossible. One of the main characteristics of these networks is that the network lifetime is highly related to the route selection. Unbalanced energy consumption is an inherent problem in WSNs characterized by the multi-hop routing and many-to-one traffic pattern. This uneven energy dissipation in many routing algorithms can cause network partition because some nodes that are part of the efficient path are drained from their battery energy quicker. To efficiently route data through transmission path from node to node and to prolong the overall lifetime of the network, In this thesis we proposed three new routing algorithms using a combination of both Fuzzy approach and A-star algorithm seeks to investigate the problems of balancing energy consumption and maximization of network lifetime for WSNs :A-Star with 3 parameters fuzzy system (A*3F), A-Star with 3 fuzzy system with 2 parameters using majority vote (A*3FMV) and A-Star with 3 fuzzy system with 2 parameters using simple additive weighting (A*3FSAW). The new methods is capable of selecting optimal routing path from the source node to the sink by favoring the highest remaining energy, minimum number of hops, lowest traffic load and energy consumption rate. We evaluate and compare the efficiency of the proposed algorithms with each other methods under the same criteria in four different topographical areas. Simulation results show that A*3PFSAW and A*3PFMV balances the energy consumption well among all sensor nodes and achieves an obvious improvement on the network lifetime that randomly scattered nodes and flat routing.

Keywords: Wireless Sensor Networks, A-Star algorithm, Fuzzy logic, Network lifetime, Multi-hop routing.

16. PaperID 31051654: Allocation Algorithm based on CAC Scheme for LTE Network (pp. 120-127)

Radhia Khdhir, LETI Laboratory, ENIS, University of Sfax, Tunisia

Kais Mnif, LETI Laboratory, ENETCOM, University of Sfax, Tunisia
Khitem Ben Ali, LETI Laboratory, ENIS, University of Sfax, Tunisia
Lotfi Kammoun, LETI Laboratory, ENIS, University of Sfax, Tunisia

Abstract — To reduce network congestion and to guarantee a certain level of Quality of Service (QoS) for service requests, Call Admission Control (CAC) as a part of Radio Resource Management (RRM) aims to accept or reject a call based on available resources. In this paper, we proposed new CAC and resources allocation schemes for Long Term Evolution (LTE). The proposed CAC scheme gives the priority of Handoff Calls (HC), without totally neglecting the requirements of a New Calls (NC). The main objective of this approach is to provide QoS and to prevent network congestion. Simulation results show that the call admission control scheme leads to increased session establishment success and resource utilization compared with existing admission control and resources allocation schemes. Moreover, the resources allocation scheme achieves a considerable gain in the system throughput and fairness.

Keywords — Call admission control; QoS; Scheduling; LTE; Uplink; Throughput.

17. PaperID 31051657: A Facebook Identical Data Detection and Deletion Algorithm (pp. 128-134)

Harshita Shukla, Dept. of Computer Engineering and Applications, National Institute of Technical Teachers' Training and Research (NITTTR), Bhopal (M.P), India.
Shailendra Singh, Senior Member IEEE, Dept. of Computer Engineering and Application, National Institute of Technical Teachers' Training and Research (NITTTR), Bhopal (M.P), India.

Abstract — Facebook is becoming very popular as millions of users are sharing their thoughts by using various data formats. The motive behind its launch was to find old friends and relatives and make new friends. All Social Networks need to meet the increasing user demands of data storage and retrieval. The Social Networks are based on cloud to deal with dynamic speed of data generation. The success of Facebook has resulted in increased user traffic and large amount of data is continuously generated by its users'. It requires novel ways of storing data and removal and removal of duplicates as much as possible while maintaining the speed of responding to a query. In this paper, an attempt is made for the identification of data duplication and its removal. Social networking sites need dynamic data management by identifying duplicate data and its deletion technique. The removal of duplicate data is necessary, not only to reduce runtime, but also to improve search accuracy and efficiency. The implementation of this method reduces the indexing time to a great extent by decreasing the collection length, resulting in the reduction of the amount of hardware required to support the system.

Keywords- Hashing; indexing; similarity checking; unique documents; detecting replicate; data duplicity; web mining; Facebook.

18. PaperID 31051660: Rule Generation for Proton Pump Inhibitor Regimen Using Learning Vector Quantization and C4.5 (pp. 135-140)

Anifuddin Azis, Department of Computer Science and Electronics, Gadjah Mada University, Yogyakarta, Indonesia
Sri Hartati, Department of Computer Science and Electronics, Gadjah Mada University, Yogyakarta, Indonesia
Edi Winarko, Department of Computer Science and Electronics, Gadjah Mada University, Yogyakarta, Indonesia
Zullies Ikawati, Department of Pharmacy, Gadjah Mada University, Yogyakarta, Indonesia

Abstract — The excessive or irrational use of drugs categorized as Proton Pump Inhibitor (PPI) was indicated in Baptis Hospital of Kediri, Indonesia. In the PPI-based drug regimen among patients with digestive disorders from December 2009 to February 2010, many cases that the PPI-based drug regimen was not in accordance with the prevailing procedures were found, i.e. the drug regimen among patients who should not be given it. In this study, a method was developed to generate the PPI-based drug regimen rule. Data on the PPI-based drug regimen were trained using Learning Vector Quantization (LVQ) algorithm. The results of LVQ were stored as new data, which were extracted into IF-THEN rule with C4.5 algorithm. Based on the test, eighteen rules were generated for the PPI-based drug regimen with an accuracy rate of 82.5% on test data.

Keywords—PPI-based drug regimen; rule generation; LVQ; C4.5

19. PaperID 31051661: APMS: Construction and Assessment of Hospital Process for Outpatients Process Analysis (pp. 141-147)

Mohammad Taha Khan, Suresh Gyan Vihar University JaipurRajasthan India, email: ertaha82@gmail.com
Dr. Shamimul Qamar, College of Computer Science, King Khalid University Abha KSA, email:
Dr. Ripu Ranjan Sinha, Suresh Gyan Vihar University Jaipur India

Abstract - Management Information Systems is the process of transforming the accumulated data into useful and helpful information systems. This paper work is on design and construction of Advanced Pathology Management System (APMS). The objectives of the APMS is to i) Well-secured login system ii) Simple and easy patient registration form iii) Better test processing system i.e scheduling for the test and tracking the reports iv) Efficient Report Management system i.e, creation, searching and verification of the required reports v) Well-defined privacy management systems. The developed APMS is tested over Urgent care hospital, New Delhi. The event logs of outpatients are accumulated from the hospital and preprocessed using process mining approaches. Performance indices such as wait time for consultation wait time for test and the aggregate time spent on the outpatient care are analyzed. Experimental results prove the efficiency of the developed Advanced Pathology Management System (APMS).

Keywords: Management Information Systems, Clinical Pathology, Report Management, Outpatients and Process mining approaches.

20. PaperID 31051666: Anonymity of Base Station in Wireless Sensor Network via Backup Base Station (pp. 148-154)

Faisal Taj (1), Shahzad Anwar (2), Muhammad Imran (1), Irshad Ullah (1)
(1) Department of Computer Science, Iqra National University, Peshawar Pakistan
(2) Institute of Mechatronics, University of Engineering and Technology Peshawar, Pakistan

Abstract - Sensor nodes covers surrounding area and report any events to a base station over multi-hop communication. The base station plays a key role in the network. The adversary, wants to disrupt network operation, would excitedly look for the base station and target it with attacks in order to inflict maximum damage. To avoid maximum damage a novel approach is proposed for boosting the anonymity of the base station. In the proposed research the numbers of base stations are increased from one to many (such as 2 to 5) in the network operation. The purpose is to divert the adversary attention about the base station and adversary considers the base station as a sensor node. Experimentation results suggest that the approach provide a backup facility in case if one of the base stations is failed due to adversary or due to energy failure. Therefore enhances network security.

Keywords – Anonymity, Base Station, Backup Base Station, Wireless Sensor Network

21. PaperID 31051668: Neural Feed Forward Fault Tolerant Backbone Tree Construction to Increase the Lifetime of Wireless Sensor Network (pp. 155-159)

K. Vimal Kumar Stephen, Mathivanan V.
Computer Science and Engineering Department, AMET University, Chennai, India.

Abstract - In the recent times, the demands of Wireless Sensor Networks (WSN) increase the challenges in terms of scalability and energy efficiency. One of the key challenges in the wireless sensor network is how to prolong the lifetime of the network. To improve the lifetime of the sensor, static and movable mobile sinks are deployed. Movable sinks are used to receive sensed data from the sensor where it is located. The static mobile sinks act as a trusted third party for computing and distributing keys between sensor nodes and the clusters. It is not necessary to chose new

cluster head often because of trusted third party sink, performs all the computations of cluster head. The energy is retained when computation is reduced in cluster head thereby increases the life time of the particular cluster. Feed forward Back propagation algorithm is proposed using adaptive learning in neural networks followed by link aware routing. This algorithm deals with fault tolerant backbone tree construction for data transmission whereas it produces optimal path for the sink to transmit data. Since the optimal path is established, the life of the sink also to be prolonged thereby increase the overall network lifetime. Result shows that the lifetime of the network is improved and energy depletion is reduced.

Keywords – Sensor Networks, mobile sink, clusters

22. PaperID 31051669: An Efficient Neural Network Model for Software Effort Estimation (pp. 160-167)

Nesa Khandoozi GholiAbad, Islamic Azad University, Gorgan branch, Iran
Sanaz Khandoozi GholiAbad, Islamic Azad University, Sari branch, Iran

Abstract — Software development effort estimation is the process of predicting the effort required to develop a software system. Estimating development effort accurately in the early stage of software life cycle plays a crucial role in effective project management. Effort estimation is a key factor for software project success, defined as delivering software of agreed quality and functionality within schedule and budget. Traditionally effort estimation has been used for planning and tracking project resources. It has become an important task. This paper proposed a neural network model for software effort estimation. This model has 3 layers. The train, validation and test data used are from COCOMO data set. Inputs and targets data randomly divided in train (60 %), validation (20%) and test (20%) group. When the number of neurons in hidden layer was 20, Number of training samples was 37, number of validation samples was 13 and number of testing samples was 13, the network has best performance. In this case, the value of training, validation and testing MSE was 0.01044, 0.0475 and 0.0375 respectively and value of training, validation and testing R was 0.9167, 0.7741 and 0.7410 respectively.

Keywords- Software Engineering, Effort Estimation, Artificial Neural Network

23. PaperID 31051674: An Efficient Approach for Digital Image Splicing Detection Using Adaptive SVM (pp. 168-173)

Gurleen Kaur, Baljit Singh Khehra,
Department of Computer Science and Engineering, Baba Banda Singh Bahadur Engg. College Fatehgarh Sahib, India

Abstract — Forgery detection is the most important task in our national judicial system and criminal investigation procedure. Today digital images have become powerful source of communication. With the advancement of technology, it becomes very easy to change the content of digital images. Due to which these images are no more taken as a proof of authenticity or legitimacy. In this paper, we deal with the widely used form of image tampering known as image composition(or image splicing).We demonstrate an effective algorithm to detect the spliced images based on illumination inconsistencies present in images. An adaptive support vector machine (a-SVM) is used to classify the given images as either genuine or forged.

Keywords—Digital image forensic, forgery detection, image splicing, Adaptive SVM.

24. PaperID 31051675: Comparison and Analysis of Image Splicing Detection Using Artificial Neural Networks (pp. 174-178)

Lovepreet Kaur, Dr. Baljit Singh Khehra
Department of CSE & IT, BBSBEC, Fatehgarh Sahib, India

Abstract — Due to advancement in technology it is easy to modify the digital images and the discovery of modified images can be the difficult task as the images are the very powerful source of communication in every field. So, one of the major issue in today's world regarding digital images is the authenticity of given images. Therefore, digital image forgery detection is a growing research field with important implication for ensuring the credibility of digital images. In this research, we proposed a credible method to detect image splicing based on illuminant color. Artificial neural network techniques are implemented as a classifier to detect the tampered images. The results describe that artificial neural network is effective to detect tampered images.

Keywords— Forgery Detection, Image splicing, Illuminant color, Artificial Neural network.

25. PaperID 31051676: Efficient Random Sampling Statistical Method to Improve Big Data Compression Ratio and Pattern Matching Techniques for Compressed Data (pp. 179-184)

Dr. Nishad P M, Department of Computer Applications, Mar Athanasios College for Advanced Studies Tiruvalla Kerala, India.
Syam Sankar, Department of Computer Science & Engineering, NSS College of Engineering Palakkad, Kerala, India

Abstract - This paper surveys various possibilities for pattern matching in compressed big data volume. Although various compression standards are available for compressing data, entire volume decompression is compelled before pattern matching, this in turn leads to increase in computational complexity as well as the space complexity. Some compressions algorithms give better compression ratio, at the same time, they are inefficient in decompression required for pattern matching. This paper evaluates the possibilities of pattern matching after compression without decoding. Also this paper experiments and proposes how the random sampling and its statistics will help to make better compression ratio in big data. The another objective of this work is to investigate the possibilities of pattern matching in big data without decoding and some of the standards are suggested based on this study and survey.

Keywords - Compression, Encoding, Decoding, Big data, compression ratio, computational complexity, space complexity, random sampling.

26. PaperID 31051686: A New Dynamic Data Replication Algorithm to Improve Execution Time in Data Grid (pp. 185-190)

*Soheila Malmoli Abbasi, * Mohammadreza Noorimehr*
Department of Computer Engineering, Khouzesan Science and Research Branch, Islamic Azad University, Ahvaz, Iran
Department of Computer, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran

Abstract — Data grids provide large-scale geographically distributed data resources for data intensive applications. These applications handle large data sets that need to be transferred and replicated among different grid sites so availability and efficient access are the most important factors affecting the performance. It is obvious that, managing the volume of data is very important. Data replication is an important technique to reduces data access time which improves the performance of the system by creating identical replicas of data files and distributing them on grid sites. In this paper, we propose a novel dynamic data replication strategy called DRPF (Dynamic Replication of Popular File), which is based on access history and file's popularity. As grid sites within a virtual organization(VO) have similar interest of files, the basic idea of DRPF is to improve locality in accesses through increasing the the number of replicas in the VO. DRPF first selects the popular files that are needed to be copied to other nodes, then tries to find the best places for placement of new replicas by taking into account parameters such as the number of demands per site for files and bandwidth between replication sites. The algorithm is simulated using a data grid simulator, OptorSim. The simulation results show that our proposed algorithm has better performance in comparison with other algorithms in terms of job execution time and effective network usage.

Keywords-Data grid; replication; popular file; placement

27. PaperID 31051687: Image Steganography Method for Concealing Secret Data into Coefficients Based on High Scalable Sub-Bands of Integer Wavelet Transform (pp. 191-197)

Yahya E. A. Al-Salhi, Songfeng Lu
School of Computer Science, Huazhong University of Science and Technology, Wuhan 430074, PR china

Abstract — In information security, an image steganography technique uses one of the most popular transforms; either a spatial domain or the frequency domain to conceal the secret information. In this paper, an image steganography system using the spatial domain technique to conceal secret information in the frequency domain is proposed to conceal secret image information in another cover image. The Integer Wavelet Transform (IWT) used to obtain high scalable sub bands for each LL, LH, HL and HH of the cover image file. Then, the steganography approach is used to conceal the secret information in the wavelet coefficients for all sub bands. The results show high quality of stego image, and the stego image is analyzed for different attacks. It is found that the technique is robust, and it can withstand the attacks. The quality of the stego image is measured by Peak Signal to Noise Ratio (PSNR), Structural Similarity Index Metric (SSIM), and Universal Image Quality Index (UIQI). The quality of extracted secret image is measured by Signal to Noise Ratio (SNR) and Squared Pearson Correlation Coefficient (SPCC).

28. PaperID 31051693: Managing and Tracking Alumni in Saudi Universities (pp. 198-204)

Dr. Amr Jadi, Department of Computer Science and Engineering, College of Computer Science and Engineering, University of Hail, Hail, Saudi Arabia

Abstract — Managing Alumni System is one of the greatest challenges in the present market of Saudi Arabia. An alumni system is a channel between different universities and labor market to deliver various services to students as per the merit and priorities. There is no constructive method in present system of Labor office to monitor job requests from the students and communicate them with potential changes of market policies. This research aims to provide an architecture building a Functional Alumni System in Saudi Universities. The loop holes of current alumni system are highlighted and a consolidated methodology is implemented to develop a unique approach for increasing challenges. To overcome these deficiencies between Alumni Systems and Labor Market, the preset research provides a runtime monitoring system based on Labor policies to attain quality and manageability. The requests placed by students, applications executed by labor office and job requests in pending can be monitored and processed with a flexible approach by using this method. In turn lot of financial wastage can be avoided by reducing the complexity between job seekers and providers by the proposed approach.

Keywords - Runtime Monitoring, Policy, Alumni System, Saudi Universities, Labor Office, Integration

29. PaperID 31051694: Secured Data Transmission in Wireless Sensor Networks (pp. 205-215)

S. Suresh, Department of Information Technology, SRM University, Kattankaluthur Campus, Kancheepuram District, TamilNadu, India
Giridhar R., Department of Information Technology, SRM University, Kattankaluthur Campus, Kancheepuram District, TamilNadu, India

Abstract — Security is one crucial requirement in Wireless Sensor network. To overcome this issue, security protocol called Didrip was developed for flat based network which allows for distributed data discovery and dissemination. But in terms of clustering approach which is most efficient one in terms of energy conservation, there are lot of security vulnerability i.e. checking the cluster head for vulnerability to the network. In addition sensor nodes joining the cluster head during user joining phase is also not secure as the nodes can be vulnerable too. These two are most vulnerable security issues which are not addressed in existing security protocol of WSN including the one mentioned which is Didrip. The above said problems for clustering approach in WSN are overcome with a Cluster-based Certificate Authority (CA) scheme which is combination of voting and Nonvoting schemes towards detecting malicious node.

We also use digital signature to sign all the nodes present in the network. These are simulated using standard network simulator ns-2 and results analysed in terms of packet delivery, network life time and energy efficiency.

Keywords - Didrip, WSN, CA, ns-2

30. PaperID 31051696: A Multi-step Method to Calculate the Equilibrium Point of the Continuous Hopfield Networks: Application to the Max-stable Problem (pp. 216-221)

Mohammed El ALAOUI (1), Karim EL MOUTAOUAKIL (2) and Mohamed ETTAOUIL (1)
(1) Modelling and Scientific Computing Laboratory, University Sidi Mohammed ben Abdellah, Fez, MOROCCO
(2) National school of applied sciences Al-Hoceima (ENSAH) BP 03, Ajdir Al-Hoceima

Abstract — The Continuous Hopfield Networks (CHN) is a neural network tools which can be used to solve many problems like auto-memory and optimization problems. The dynamics of the CHN is described by differential equations system which is hard to solve analytically. That is why, the researchers use the Euler Cauchy method to calculate the CHN equilibrium point. Unfortunately, this method suffers from several problems, especially quality of the decision for a large step, sensibility to the slope function parameters and to the initial conditions. In this work, we use the well-known multi-step numerical method called Adams–Bashforth method, which is strong in terms of stability and performance, to calculate the equilibrium point of the CHN associated with the max stable problem. This method introduces an intermediary step to improve the Euler Cauchy method precision. The experimental results show that the (CHN+Adams-Bashforth) method produce a large max stable sets in comparison with the (CHN+Euler-Cauchy) method.

Keywords: - Continuous Hopfield Networks, Euler Cauchy method, Adams–Bashforth method, max-stable problem.

31. PaperID 31051699: An Event Grouping Based Algorithm for University Course Timetabling Problem (pp. 222-229)

Velin Kralev, Radoslava Kraleva, Borislav Yurukov,
Department of Informatics, South West University "Neofit Rilski", Blagoevgrad, Bulgaria

Abstract — This paper presents the study of an event grouping based algorithm for a university course timetabling problem. Several publications which discuss the problem and some approaches for its solution are analyzed. The grouping of events in groups with an equal number of events in each group is not applicable to all input data sets. For this reason, a universal approach to all possible groupings of events in commensurate in size groups is proposed here. Also, an implementation of an algorithm based on this approach is presented. The methodology, conditions and the objectives of the experiment are described. The experimental results are analyzed and the ensuing conclusions are stated. The future guidelines for further research are formulated.

Keywords – university course timetabling problem; heuristic; event grouping algorithm

32. PaperID 300416114: Digital Image Watermarking Using DCT and DWT to Improve Robustness (pp. 230-234)

Priyanka Rani, Anupam Singh, Computer Science and Engineering, Shri Ramswaroop Memorial University, Lucknow, India
Avinash Kumar Singh, Computer Science and Engineering, Shri Ramswaroop Memorial University, Lucknow, India

Abstract — Watermarking is the concept that provides protection in digital multimedia. This paper uses Discrete Wavelet Transform (DWT), Singular Value Decomposition (SVD) and Discrete Cosine Transform (DCT) concept for watermarking and extraction purpose. In result analysis we analyze extracted image from watermarked image after applying different attacks (like rotation, Gaussian noise, average filter attack, low pass filter, high pass filter, salt and

pepper, Histogram Equalization etc). We find that this concept is robust against these types of attacks and provide high security.

Keywords- Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Singular Value Decomposition (SVD), Cover Image, Watermark Message.

33. PaperID 310316102: A New Efficient two tire secure protocol (pp. 235-240)

Rehan Ullah, IT Department Hazara University, Mansehra, Pakistan
Noor ul Amin, IT Department Hazara University, Mansehra, Pakistan
Faisal Bahadur, IT Department Hazara University, Mansehra, Pakistan
Abdul Hakeem, IT Department Hazara University, Mansehra, Pakistan
Insaf Ullah, IT Department Hazara University, Mansehra, Pakistan

Abstract — Signcrypion is a cryptographic method in which signature and encryption apply on message in a single step. On other hand image steganography is a strongest technique for hiding data or information. Therefore Communication through insecure channel is challengeable task for an organization. Recently two tier security gain popularity because most of the business organizations wants maximum security of data/information. In this paper we design a new scheme using cryptographic and stenographic techniques at once on the basis of image steganography and elliptic curve cryptography. In proposed design scheme we use both of the steganography as well as cryptography. The cryptographic technique encrypts the data by using Elliptic curve cryptography in such a manner that third party not understands the original message contents. Stenographic technique is used to hide the text in image and then we take hash as well as signature. It also assures the security properties like message confidentiality, message integrity, message non repudiation and also message authentication.

Keywords-component Cryptography, Steganography, Signcrypion, Elliptic curve cryptography.

34. PaperID 310516111: Formal Model of Smart Traffic Monitoring and Guidance System (pp. 241-252)

Umber Noureen Abbas, Farhan Ullah, Nazir Ahmad Zafar
Department of Computer Science, COMSATS Institute of Information Technology Sahiwal COMSATS Road off GT road, Sahiwal 57000, Pakistan

Abstract — Emergency Services Rescue 1122 and Smart Sticker components of our proposed Smart traffic monitoring and guidance system model are presented in this paper to provide smart emergency services and to identify vehicles to develop advanced transportation system. It involves the Wireless Sensors and actors to communicate with the system. The proposed components require fewer resources in terms of sensors and actors. Further, Sensors component identifies vehicles through Smart Stickers and it is readable through sensors from its barcode and barcode consists of vehicles details in terms of vehicles registration, model, engine and color. Secondly, Emergency Services Rescue 1122 component provides emergency services as it locates the vehicles through sensors and informs the local authority for providing emergency services. Third, violation of rules detects intruders on roads to provide smooth flow of traffic. Fourth, to avoid congestion, traffic signals are configured and communicated with sensors to update the system if congestion occurs. The proposed components of our model are implemented by developing formal specification using VDM-SL. VDM-SL is a formal specification language used for analysis of complex systems. The developed specification is validated, verified and analyzed using VDM-SL Toolbox.

35. PaperID 310516113: Anonymous and Secure Routing Protocol for Multi-hop Cellular Networks (pp. 253-258)

Salwa Othmen (1), Faouzi Zarai (1), Aymen Belghith (2), Lotfi Kamoun (1)
(1) LETI laboratory, University of Sfax, Tunisia
(2) Saudi Electronic University (SEU), Computer Science Departement, Saudi Arabia

Abstract — In single cellular networks, the mobile stations cannot communicate directly with each other. All communications are relayed through the base stations. Such topology suffers from many limitations such as congestion problem when a large number of users are communicating in the same time to a base station. In this context, the device-to-device communications have been proposed to overcome the limitations of the conventional cellular architecture. Indeed, a mobile station can allow two nearby stations to communicate with each other without involving a base station. However, security becomes an important challenge that must be taken into consideration as the mobile stations participate in routing data between each other. In this paper, we propose a secure routing protocol for Multi-hop Cellular Networks (MCNs). Our goal is to discover a secure and short route between the source and the destination. To evaluate this proposed protocol, we perform some simulations using Network Simulator (NS-2). The simulation results show that it provides acceptable performance in terms of throughput and routing overhead as comparing with Secure Ad hoc on demand Distance Vector (SAODV).

Keywords-component; single cellular networks, base stations, Device-to-device, secure routing protocol, MCNs, NS-2;

36. PaperID 310516118: Performance Analysis of Heterogeneous Data Normalization with a New Privacy Metric (pp. 259-264)

J. Hyma (†), PVGD Prasad Reddy (††), and A. Damodaram (†††)
(†) Department of CSE,GIT, GITAM University, Visakhapatnam, INDIA
(††) (†) Department of CS&SE, AU College of Engineering, Andhra University, Visakhapatnam, INDIA
(†††) Department of CSE, Sri Venkateswara University, Tirupathy, INDIA

Abstract - Investigation on privacy preserving data mining is in extensive need to the present day technological situation. Storage of the data and its usage through various computational processes is becoming very easy and efficient. At the other end the primary concern or sometimes can be termed as limitation to this extensive data analysis is privacy. There are existing privacy preserving techniques that solve this problem and also guarantee privacy as well as data utility. But these techniques have to be updated in parallel to the expansion of digital technology. In view of this, the part of research in this paper analyses various normalization techniques with heterogeneous data distortion. The experimental consideration is done with the comparison of various statistical measures on the distorted data and their preservation with respect to the original data. We evaluated the performance of heterogeneous data distortion with three types of transformations namely Min-Max Normalization, Z-Score Normalization and Decimal Scaling. The performance is evaluated with various data distortion measures and privacy measures.

Keywords: Privacy Preserving Data Mining (PPDM), Data Normalization, Privacy, Data utility.

37. PaperID 310516121: Image Compression using Clustering Algorithms (pp. 265-268)

Lale Fathi Ajirlou, Department of Computer, Germi Branch, Islamic Azad University, Germi, Iran
Seyed Naser Razavi, Department of Computer, Tabriz Branch, Islamic Azad University, Tabriz, Iran

Abstract — There is a correlation between pixels in each image so that each pixel value of adjacent pixels can be guessed. By removing these dependencies can be compressed images. Our goal is to reduce the amount of compressed image data needed to display the digital images and therefore reduce the cost of transmission and storage. Compression has a key role in many important applications. These applications include image database, transmission of images, remote sensing, medical imaging, military and space equipment remote control and so on. In addition to the compression, image coding, there's talk. That after quantization matrix should be coded range of conversions. In reconstruction after decoding to achieve our desired image obtained with the difference that the picture is far less than the original image. What we've done in this thesis using a fractal method utilizes a Kohonen neural networks and clustering to increase the compression ratio and reduction coding and decoding the image. We have implemented three methods based on fractal coding. The first method is simple fractal coding. In the second method to create the codebook of multiple tree fractal coding is used. In the second method of vector quantization LBG algorithm for

Kohonen neural network-based clustering algorithm and code book for coding image is used. Results in the second method show faster encoding. The method is simple fractal compression rate is higher than other methods.

Keyword: image compression; clustering; vector quantization

38. PaperID 310516122: A Joint Duty Cycle and Optimal Energy Adaptation Algorithm for the Body Area Sensor Networks (pp. 269-274)

Ali Raza, Dept. of computer science, City University of Science & Information Technology, Peshawar, Pakistan
Arshad Farhad, Dept. of computer science, COMSATS, Sahiwal, Pakistan
Wajid Ullah Khan, Dept. of computing, Abasyn University, Peshawar, Pakistan
Muhammad Arif, Dept. of computer science, City University of Science & Information Technology, Peshawar, Pakistan

Abstract — IEEE 802.15.4 standard is widely adapted for Body Area Sensor Networks (BANs) due to its low duty cycle and low power operation. However, IEEE 802.15.4 recommends the use of fixed duty cycle operation which results in high energy consumption and end-to-end delay. Therefore, an efficient algorithm is needed to adapt duty cycle operation to overcome the end-to-end delay and energy consumption. In this paper, we propose a Joint Duty Cycle algorithm (JDCA) for the BAN to enhance the network lifetime, throughput and decrease the end-to-end delay. Dynamic duty cycle can be adapted by the two MAC parameters: Beacon Order (BO) and Super frame Order (SO). However, these parameters are set by the network administrator before the network deployment. During simulation, JDCA algorithm is capable of adapting dynamic duty cycle at run time based on traffic load. Furthermore, simulation results shows enhanced network lifetime, network throughput and less end-to-end delay when compared with IEEE 802.15.4.

Index Terms — Dynamic duty cycle, IEEE 802.15.4, Body area sensor networks, Wireless personal area network.

39. PaperID 310516124: Performance Evaluation of High Performance Data Transfer in Grid Environment over Broadband Hybrid Satellite Constellation Communication System (pp. 275-279)

Anupon Boriboon, Vincent Mary School of Science and Technology, Assumption University of Thailand, Bangkok, Thailand
Siriwhaddhanah Pongpadpinit, Department of Business Information System, Martin de Tours School of Management and Economics, Assumption University of Thailand, Bangkok, Thailand

Abstract — This paper presents the evaluation performance of broadband hybrid satellite constellation communication system (BHSCCS) networks which provides high performance data transfer in grid network environment based on TCP protocols. The evaluated hybrid satellite network uses the COMMStellationTM constellation topology on lower orbital. We adopt the GridFTP to improve network performance. GridFTP is a high-performance, reliable data transfer protocol optimized for high-speed Internet to suitable WAN networks. The simulation results show the network performance of GridFTP which different AQMs, TCPs, PERs, over BHSCCS networks.

Keywords: COMMStellationTM; GridFTP; Hybrid Satellite; Queue; TCP

40. PaperID 310516127: A Lasso-LTS Method for DNA Sequence Classification Based on Beta Wavelet Networks (pp. 280-292)

Abdesselem DAKHLI, Department of Computer Science, REGIM, University of Gabes, Tunisia
Wajdi BELLIL, Department of Computer Science, REGIM University of Gafsa 2110 Gafsa, Tunisia
Chokri BEN AMAR, Department of Computer Science, REGIM University of Sfax 3018,Sfax, Tunisia

Abstract — Wavelet Neural Network (WNN) is attracting interest in field of classification system, because they are universal approximations, particularly due to rapid and accurate representation of nonlinear dynamic systems. The satisfying performance of the WNN depends on an appropriate determination of the Wavelet Neural Network structure. In this paper we provide a new method to solve this problem based on the Least Absolute Shrinkage and Selection Operator (LASSO). At first, the scale of WNN is managed by using the time-frequency locality of wavelet. Furthermore, the unconstrained optimization problem (LASSO) is used to solve the structure and learning of the WNN. This optimization problem can be solved efficiently using the iteratively reweighted least squares (IRLS) and the Least Trimmed Square (LTS) methods to enhance the ineffectiveness; they are applied to train the wavelet neural network. The advantage of the method lies in the oracle properly of the LASSO can guarantee the optimal structure of the WNN. The proposed method has been able to optimize the wavelet neural network and this method is able to classify the DNA sequences. Our goal is to construct predictive models that are highly accurate. In fact, the proposed method permits to avoid the complex problem of form and structure in different clusters of organisms. The empirical results and their classification performances are compared with other methods. We compared the WNN-Lasso model with the other five alignment-free models, i.e., k-tuple, DMK, TSM, AMI, and CV, on several large-scale DNA datasets on the DNA classifying application by means of the K-means method. The experimental results have shown that the WNN-Lasso model outperformed the other models in terms of both the classifying results and the running time. Evenly, in this study, we present our approach consists of three phases. The first one, which is called transformation, is composed of two sub steps; binary codification of the DNA sequences and the Signal Processing of the DNA sequences. The second phase step is the approximation; it is empowered by the use of the Multi Library Wavelet Neural Networks (MLWNN). Finally, the third section, which is the classification of the DNA sequences, is realized by applying the algorithm of k-means classification.

Index Terms— LASSO, LTS, Wavelet Neural Networks, DNA sequences, MLWNN, IRLS.

41. PaperID 310516129: Sindhi Morphological Analysis: An Algorithm for Sindhi Word Segmentation into Morphemes (pp. 293-302)

Waqar Ali Narejo, Javed Ahmed Mahar, Shahid Ali Mahar, Farhan Ali Surahio, Awais Khan Jumani
Department of Computer Science, Shah Abdul Latif University, Khairpur Mir's, Sindh, Pakistan

Abstract- Morphological analysis is the process of constructing and deconstructing the words of a language, the process is based on the basic grammatical units which are stem, prefixes, suffixes and infixes. Sindhi is rich in morphological features with a great variety of affixes. The problem for Sindhi to come into computerization is the large number of variants in its morphology. This complexity is created due to different positions of prefixes, suffixes and stems in the words. The automatic word segmentation system normally faces such embedded hurdles in Sindhi language. An algorithm is required with a capability of dealing with such issues for the segmentation of Sindhi words. In this paper, an algorithm is designed and implemented to resolve the problem of segmenting Sindhi complex and compound words into possible morphemes. The developed words segmentation system has been tested on a list of 109 compound words, 179 prefix words, 1343 suffix words and 50 prefix-suffix words. The cumulative segmentation error rate of 5.02% is calculated. This system can also be used as pre-requisite in various Sindhi language and speech processing applications.

Keywords — Sindhi Morphology; Morphological Analysis; Word Segmentation; Morphemes

42. PaperID 310516130: A New Secret Sharing Scheme Using Rational Interpolation (pp. 303-307)

Ali Nakhaei Amroudi, Department of Mathematics and Cryptography, Malek Ashtar University of Technology, Isfahan, Iran
Ali Zaghian, Department of Mathematics and Cryptography, Malek Ashtar University of Technology, Isfahan, Iran,
Mahdi Sajadieh, Department of Electrical Engineering, Islamic Azad University, Isfahan (Khorasgan) Branch, Isfahan, Iran

Abstract — Most of the existing secret sharing schemes are based on polynomial interpolation. In other word, they use polynomial functions in their schemes. In this paper, we solve the problem of creating a secret sharing scheme based on rational interpolations. We show that if * support points have the same width then the rational interpolation of the support points, which is called ( )( ), has pole points. Finally, we give an example for the accuracy of the proposed scheme.

Keywords-component; Secret Sharing Scheme; Shamir's Scheme; Polynomial Interpolation; Rational Interpolation, Pole Points.

43. PaperID 310516133: A Novel Face Recognition System based on Skin Detection, HMM and LBP (pp. 308-316)

Mejda Chihaoui, Akram Elkefi, Wajdi Bellil, Chokri Ben Amar
REGIM: Research Groups on Intelligent Machines, University of Sfax
National School of Engineers (ENIS), Sfax, 3038, Tunisia.

Abstract — Although there are various biometric techniques, like fingerprints, iris scan as well as hand geometry, the most efficient and widely-used one is face recognition because it is inexpensive, non-intrusive and natural. In our paper, we present an approach aiming at implementing a full architecture which represents an efficient system of face recognition. For this, an attempt is proposed for each system stage. At the beginning, we develop a novel approach to detect faces existing in 2D color image. This approach focuses mainly on how to implement a selection of skin color before using neural networks and Gabor filters. This approach represents an improvement of existing approach especially because it aims to minimize the computation time. Indeed, the skin detection step avoids wrong detection and to help the system detect the face in the right areas and minimize the research time and subsequently the Gabor filter will be applied only on the localized skin space. Later, the face features obtained by the Gabor filter represent the input of the neural network classifier to decide whether an input image pixel is a face pixel or not. For 2D face recognition, we propose likewise a novel approach that we call HMMLBP (a combination of the two tools Hidden Markov Models HMM and Local Binary Pattern LBP). It allows classifying a given 2D face image through utilizing an LBP tool to extract features. In order to validate our whole system performance, we show experimental results obtained when applying our proposed algorithm on benchmark face databases, respectively AT&T, Yale and Feret.

44. PaperID 310516134: Energy Efficiency Techniques in Cloud Computing (pp. 317-323)

Altaf Ur Rahman, Fiaz Gul Khan, Waqas Jadoon
Department Computer Science, COMSATS Institute of Information Technology, University Road Tobe Camp, Abbottabad

Abstract — Cloud computing gaining popularity at enormous rate since from its emergence. CC changed the way that computing services are provided. On demand platform (PaaS), infrastructure as a service (Iaas) and software (SaaS) as a service through internet. Consumer use third party services instead of building his own infrastructure which need up-front investment and expertise. Cloud computing becoming popular for unlimited computing power, availability, nice pricing, on demand services and quality of service. For availability and computing power the service provider expands their resource capacity to handle user requirements. This expansion in resources capacity lead to high energy demand. Two big issues for cloud computing is energy demand and security/privacy requirements. In this survey we will give a review on the latest techniques for energy efficiency in cloud computing. The main focus is on software base energy efficiency techniques in which we will explain the workload consolidation and resource management in detail.

Index Terms — cloud computing, data center, energy efficiency techniques.

45. PaperID 310516135: Service Level Agreement in Cloud Computing: A Survey (pp. 324-330)

Usman Wazir, Dr. Fiaz Gul Khan, Dr. Sajid Shah
Department of Computer Science, COMSATS Institute of Information Technology, University Road Tobe Camp, Abbottabad, Pakistan

Abstract — Cloud computing provides distributed resources to the users globally. Cloud computing contains a scalable architecture which provides on-demand services to the organizations in different domains. However, there are multiple challenges exists in the cloud services. Different techniques has been proposed for different kind of challenges exists in the cloud services. This paper reviews the different models proposed for SLA in cloud computing, to overcome on the challenges exists in SLA. Challenges related to Performance, Customer Level Satisfaction, Security, Profit and SLA Violation. We discuss SLA architecture in cloud computing. Then we discuss existing models proposed for SLA in different cloud service models like SaaS, PaaS and IaaS. In next section, we discuss the advantages and limitations of current models with the help of tables. In the last section, we summarize and provide conclusion.

Index Terms— Service Level Agreement (SLA), Cloud Computing.

46. PaperID 310516136: Blind Watermarking Algorithm for 3D Multiresolution Meshes based on Spiral Scanning Method (pp. 331-342)

Ikbel Sayahi, Akram Elkefi, Chokri Ben Amar
REGIM-Lab.: REsearch Groups in Intelligent Machines, University of Sfax, ENIS, Sfax,3038 Tunisia

Abstract — 3D mesh is a new data type appeared in the last decades. Since its emergence, it has been used in several areas which raise major security problems. As a solution, we propose a blind watermarking algorithm for 3D meshes. For doing spiral scanning method decomposes the mesh into GOTs (a Group of Triangles). At each time, only one GOT will be uploaded into memory. It undergoes a wavelet transform to generate vector of wavelet coefficients. This latter undergoes modulation then embedding steps using data coded with BCH code. Once watermarked, the next GOT will be uploaded. This process stopped when the entire mesh is watermarked. Experimental tests show that the quality of meshes is kept despite the high insertion rate and that memory consumption is reduced. As for robustness, our algorithm overcomes the following attacks: translation, rotation, smoothing, uniform scaling, coordinate quantization, noise addition, simplification and compression.

Index Terms — Digital watermarking, 3D meshes, Multiresolution, Wavelet transform, Spiral scanning, Attacks, Compression.

47. PaperID 310516141: Towards the Development of an Efficient and Cost Effective Intelligent Home System Based on the Internet of Things (pp. 343-350)

A. Imtar Chaudary, Dept. of Computer Science, CIIT, Sahiwal, Pakistan
Muhammad Usman, Dept. of Computer Science, CIIT, Sahiwal, Pakistan
Arshad Farhad, Dept. of Computer Science, CIIT, Sahiwal, Pakistan
Wajid Ullah Khan, Dept. of Computing and Technology, Abasyn University, Peshawar, Pakistan

Abstract — Internet of Things (IoT) is an emerging technology which is covering everyday things from industrial machinery to consumer goods in order to exchange information and complete tasks while involved in other work. IoT based smart home automation system is a system that uses PCs, mobile phones or remote devices to control basic operations for home automatically from anyplace around the world using internet. The proposed intelligent home automation system differs from existing systems as it allows the user to operate the system from anywhere around the world by using internet connection along with intelligent nodes that can take decisions according to the environmental conditions. We implemented a home automation system using sensor nodes that are directly connected to Arduino microcontrollers. Microcontroller is programmed so that it can perform some basic operations on the basis of sensors data. e.g. fan is controlled on basis of temperature value and light is controlled on the basis of occurrence of motion in the room etc. Furthermore Arduino board is connected to the internet using Wi-Fi module. An extra feature this system provides is to monitor power consumption of different home appliances. The designed system provides the

user remote control of numerous appliances locally as well as outside the home. This designed system is expandable, allowing multiple devices to be controlled. The objective of the proposed system is to provide a low cost and efficient solution for home automation system by using IoT. Results show that the proposed system is able to handle all controlling and monitoring of home.

Keywords—Internet of Things (IoT), Wireless Sensor Network, Home Automation System, Energy Monitoring.

48. PaperID 310516142: A Threshold-Based Predictive Scheme for Mobile Subscribers in Publish/Subscribe Systems (pp. 351-357)

Fatma Abdennadher, Maher Ben Jemaa
National School of Engineers of Sfax, University of Sfax, ReDCAD Laboratory, B.P.1173, 3038 Sfax, Tunisia

Abstract — In this paper, we present our strategy adopted to deal with the mobility into publish/subscribe. Specifically, we focus on the management of the mobile users from one broker to another. In fact, the topic of mobility into publish/subscribe systems may cause many problems such as the increasing of the traffic into the network and the messages loss. To overcome these problems, we have created a selective scheme on the basis of an accurate selection. In fact, a threshold value is devoted to be the criterion for the selection of caching points. On the basis of this principle, we apply various network settings to explore the effectiveness of our approach. Hence, we extract the improvement of our approach on the messages loss, the caching cost and the propagation cost in function of buffer size, publication rate, period of disconnection and connect time.

Keywords-Distributed Networks; Mobile Computing; Publish/-Subscribe; Prediction Management; Performance Efficiency.

49. PaperID 310516147: A Novel Protocol Stack for Improving QoS in Vehicular Networks (pp. 358-367)

Mohammadreza Pourkiani, Department of Information Technology Science and Research Branch, Islamic Azad University, Tehran, Iran
Sam Jabbehdari, Department of Computer Engineering, Tehran North Branch, Islamic Azad University, Tehran, Iran
Ahmad Khademzadeh, Department of National and International Cooperation, Iran Telecommunication Research Center, Tehran, Iran

Abstract — Intelligent Transportation Systems are defined as those systems utilizing synergistic technologies and systems engineering concepts to develop and improve transportation systems of all kinds. Vehicular Ad-hoc Network (VANETs) which is an application of Mobile Ad-hoc Networks (MANETs) play an important role in ITS and emerged to provide Vehicle to Vehicle, Vehicle to Roadside and Vehicle to Infrastructure communications, aiming to improve safety on roads, exchange data between vehicles and provide different services to the users. According to special characteristics of VANETs like bandwidth limitation, high mobility, signal fading and real-time data communications, QoS provisioning in these networks is a challenging task. In this paper, we introduce an architecture for vehicular networks and a protocol stack which aims to reduce the processing overhead, make routing easier and provide Quality of Service in vehicular networks. Finally, after designing protocols and headers of the mentioned protocol stack, we will simulate our proposed idea in a vehicular environment and after simulation process, we will compare the achieved results with another scenario in which regular TCP/IP protocols are used.

Keywords-component; VANETs; ITS; QoS; Protocol Stack

50. PaperID 310516149: Performance Analysis of VoIP over IPV4, IPv6 and 6-to-4 Tunneling Networks (pp. 368-372)

Muhammad Fawad (1), Syed Irfan Ullah (2), Haseena Noureen (3), Arbab Wajid Khan (4), Zar Khitab (5), Shahab Khan (6), Abdus Salam (7), Muazzam A. Khan (8)
(1,2,4,6,7) Department of Computing & Technology, Abasyn University Peshawar KP Pakistan
(3) Faculty of Computer Science & IT, University of Malakand Dir(L) Malakand KP Pakistan
(5) Faculty of Electrical Engineering, APCOMS Rawalpindi, Pakistan
(8) NUST College of EME, National University of Sciences & Technology, Islamabad Pakistan

Abstract — Transition from IPv4 to IPv6 is a cumbrous process because of their irreconcilability with each other and coexists during the transition period. This work examines the behavior of transition mechanisms that involve communication among IPv4 and IPv6 in various scenarios and traffic conditions. A network analyst faces variable traffic and data rates at different nodes in such a heterogeneous network, that requires more attention to make it able to work with stable network flow and data rate. We analyse an end-to-end delay of VOIP data packets in IPv4 and IPv6 homogeneous and heterogeneous networks using 6 to 4 tunneling techniques. This work shows that IPv6 has better performance than IPv4 and IPv6-to-IPv4 tunneling. The tunneling technique improves the network throughput and queuing delay over the intermediate nodes of the heterogeneous network.

Keywords: IPv4, IPv6, VoIP, 6- to-4 tunneling, DSTM

51. PaperID 310516151: Investigation of Collusion Attack Detection in Android Smartphones (pp. 373-379)

M. Kireet, Dept of CSE, JNTUH, Hyderabad, India
Dr. Meda Sreenivasa Rao, JNTUSIT, Hyderabad

Abstract — Today as Android is used by majority of the smartphone users it has become one of the effortless platform for the malware-writers to introduce their malicious activities into smartphone world through this android mobile applications. The main loophole in Android applications is permission based security control. The User unawareness of accepting every permission as a mandatory requirement by an app is making more and more convenient for the hackers to extract the users' private data. In this paper we have analysed all the leakages which are done by using permissions required by an app. We carefully made an investigation to detect collusion attacks .We analyzed the present detection methods of inter-permission leaks especially on Collusion attacks and mentioned the areas where the enhancements are needed with limitations that existed in present detection methods.

Keywords - Collusion attacks, inter-permission leaks

52. PaperID 310516152: A Hybrid Machine Learning Model for Selecting Suitable Requirements Elicitation Techniques (pp. 380-391)

Nagy Ramadan Darwish, Department of Computer and Information Science, Institute of Statistical Studies and Research, Cairo University, Egypt
Ahmed Abdelaziz Mohamed, Department of Information Systems, Higher Technological Institute, Cairo, Egypt
Abdelghany Salah Abdelghany, Department of Information Systems, Higher Technological Institute, Cairo, Egypt

Abstract — Requirements elicitation is the first and the most critical phase of Requirements Engineering (RE). Many techniques have been proposed to support the elicitation process. Each technique has its strengths and weaknesses. This variety makes the selection of technique or combination of techniques for a specific project a difficult task. Mostly techniques are selected based on personal preferences rather than on attributes of project, technique, and stakeholders. In this paper, the researchers propose a three-component approach for elicitation techniques selection. First, a literature review is conducted to identify the attributes affecting techniques selection and common elicitation techniques. Second, a multiple regression model is built to analyze these attributes in order to find the critical attributes influencing techniques selection. Finally, an Artificial Neural Network (ANN) based model for selecting adequate elicitation techniques for a given project is proposed. The ANN model helps reduce the human involvements in this process. It was implemented using Neural Network Fitting Tool in MATLAB. The network has accuracy of 81%. The ANN model was empirically validated by conducting a case study in a software company.

Keywords: Requirements Engineering, Requirements Elicitation, Multiple Regression Analysis, Neural Network.

53. PaperID 310516155: A Comparison of Proxy Re-Encryption Schemes – A Survey (pp. 392-397)

Anum Khurshid, Fiaz Gul Khan, Abdul Nasir Khan
Department of Computer Science, COMSATS Institute of /information Technology, Abbottabad, Pakistan

Abstract — Proxy Re-Encryption has been used since the need for forwarding an encrypted message to a party for whom it was not encrypted was highlighted in the form of delegation rights by Blaise, Bleumer and Strauss. Various Proxy Re-Encryption schemes have been introduced till today mainly focusing on demonstrating features like transitivity and collusion-resistance to ensure minimal trust on the proxy and maximum key-privacy. This survey highlights some major schemes introduced, classifies them based on their directionality, brings to light their major advantages and disadvantages, and provides a detailed comparative study based on the key features a Proxy Re-Encryption Scheme must possess in order for its widespread.

Index words— bilinear maps, CCA secure, collusion resistance, CPA secure, delegation rights, Deffie-Hellman key exchange, DBDH assumptions, Proxy Re-Encryption; transitivity.

54. PaperID 310516163: Energy Efficient Routing Protocols in Wireless Sensor Networks: A Survey (pp. 398-406)

Owais Khan, Fiaz Gul Khan, Babar Nazir, Usman Wazir
Department of Computer Science, COMSATS Institute of Information Technology, University Road Tobe Camp, Abbottabad, Pakistan

Abstract — WSN is an evolving technology since last ten years. As wireless nodes work have less power supply in the form of a battery, it is necessary for the nodes to work for maximum time. Different techniques are adopted to achieve better energy optimization. This paper presents a survey on energy efficient routing techniques, which will help in understanding the factors which affect energy efficiency and other performance parameters and will help to analyse the techniques for further optimizations.

Index Terms — Wireless Sensor Networks, Energy optimization, Topology.

55. PaperID 310516166: Improved Face Recognition Rate Using Face Partitioning in Eigen and Fisher Feature Based Algorithms (pp. 407-417)

Harihara Santosh Dadi, Gopala Krishna Mohan Pillutla

Abstract — Face partitioning technique is presented in this paper. Instead of directly giving the face to the face recognition system, first the face is partitioned in to different face parts using face partitioning technique. The face parts are namely mouth, left eye, right eye, head, eye pair and nose. Eigen and Fisher features based algorithms are considered for experimental purpose. These face part features are given to the SVD classifiers individually. The outputs of the classifiers are again given to the decision making algorithm. Based on the maximum likely hood principle, this decision making algorithm outputs a face. ORL data base is used for evaluating the performance of this new technique. The first two faces of all the 40 people in the data base are considered for testing and the remaining eight faces are used for training purpose. Results are separately calculated with and without face partitioning technique. Results show that face recognition rate is increased by using the combination of face partitioning technique and basic face recognition algorithm. The new algorithm is also verified on 8 different data sets. Experimental results show that this face partitioning is improving the face recognition rate both Eigen and Fisher feature based algorithms.

Index Terms—Face Partitioning, Facial features, Recognition engine, Support Vector Machine, Decision making algorithm.

56. PaperID 310516168: Elastic Extension Tables for Multi-tenant Cloud Applications (pp. 418-431)

Haitham Yaish (1, 2, 3), Madhu Goyal (1, 2), George Feuerlicht (2, 4)
(1) Centre for Quantum Computation & Intelligent Systems
(2) Faculty of Engineering and Information Technology,University of Technology, Sydney
P.O. Box 123, Broadway NSW 2007, Australia
(3) Faculty of Engineering, American University of the Middle East, Kuwait
(4) Faculty of Information Technology, University of Economics, Prague, Czech Republic

Abstract — Software as a service (SaaS) is a Cloud Computing service model that exploits economies of scale for SaaS service providers by offering a single configurable software and computing environment for multiple tenants. This contemporary multi-tenant service requires a multi-tenant database that accommodates data for multiple tenants using a single database schema. In general, traditional Relational Database Management Systems (RDBMS) do not support multi-tenancy and require schema extensions to provide multi-tenant capabilities. This paper proposes a multi-tenant database schema called Elastic Extension Tables (EET), which is highly flexible in enabling the creation of database schemas for multiple tenants by extending a preexisting business domain database, or by creating tenant business domain database from the scratch at runtime. The empirical results presented in this paper indicate that the EET schema has potential to be used for implementing multi-tenant databases for multi-tenant SaaS applications.

Index Terms— Cloud Computing, Software as a Service, Multi-tenancy, Elastic Extension Tables, Multi-tenant Database.

57. PaperID 310516174: Triangle Area Based MCA Technique and Anomaly Based Detection Technique for Detecting DOS Attacks (pp. 432-440)

Varsharani Dudhande, Sharmila Wagh
Modern Education Society's College of Engg., Computer Department, Savitribai Phule Pune University, India

Abstract — The availability of network services are being menaced by the increasing number of Denial-of-Service (DoS) attacks. The availability of such interconnected systems is severally degraded by increasing number of DOS attacks. Denial-of-Service (DoS) attacks cause serious impact on these computing systems such as router, host or entire network. DoS attack detected using Multivariate Correlation Analysis (MCA) technique. Multivariate correlation analysis employs for accurate network traffic characterization by extracting the geometrical correlations between network traffic features. The proposed system uses the Multivariate Correlation Analysis (MCA) technique for accurate characterization also uses the anomaly based detection technique in attack recognition. Anomaly based detection makes system capable of detecting seen and unseen attacks. Moreover, a triangle area based technique is planned to reinforce and increases performance of MCA. The impact of each non-normalized information and normalized information on the performance of the proposed detection system is tested.

Keywords — Denial- of- Service attack, network traffic characterization, multivariate correlations, triangle area.

58. PaperID 310516176: Proposed Hybrid Model to Detect and Prevent SQL Injection (pp. 441-448)

Mrs. Teresa K. George, Dr. Rekha James, Dr. Poulose Jacob
Dept. of Computer Science, CUSAT, Cochin

Abstract - SQL Injection vulnerability takes advantages of the poorly coded web application and exploits the sensitive and critical information stored in an application's database by compromising the authentication logic of the database server. In Most of the web applications user inputs in the dynamic web pages are the vulnerable points for SQL

injection attack. A Single detection tool cannot handle the sophisticated injection attacks by the intelligent hackers. The proposed hybrid model with SQLI-Rejuvenator on an Application Program Interface is tested and proved as an efficient technique to detect and prevent SQL injection. In this architecture, the malicious queries are blocked and an alert message is generated if the injection is detected. Only the benign query is allowed to access the data from the backend database server. The Unique identity created by the template creator application, the Rejuvenator module and evaluation engine are significant features of the proposed model to prevent the Injection attack and can facilitate better availability of the application.

Keywords – Authentication; Injection; Vulnerability; Hackers; Detection; Rejuvenation;

59. PaperID 310516179: Hand Gesture Recognition System (pp. 449-453)

ABBES Zeineb, CHIBANI Chaala
Dept of industrial computing. Higher Institute of Computer and Multimedia, Gabes, Tunisia
Tarek Frikha, Abir Hadriche
CES Lab, REGIM Lab, Sfax, Tunisia

Abstract - In this article, we will propose a real-time human hand gesture recognition system which will perform translations from the sign language to the common French language. The processes is composed by three basic steps: The detection and extraction of the hand pattern characteristics during the image stream acquisition, which is obtained from an integrated camera. The analysis process, in which the obtained characteristics are classified as either a recognized sign language gesture or an unclassified hand movement. Preset characteristics of each effective hand gesture are stored locally. The message-assembling phase: at the end of cycle of each iteration of the two previous steps, the obtained result is either neglected or concatenated with the assembled message so far. The message is then displayed.

Keywords: human-machine communication, gestural interaction, French sign language, linked gesture recognition.

60. PaperID 310516180: An Optimization Technique for Brain Tumor Recognition (pp. 454-464)

Dr. D. Rajya Lakshmi, CSE Dept in JNTUK Vizianagaram
Shaik Salma Begum, CSE Dept in Gudlavalleru Engineering College

Abstract - In this paper, we have proposed a robust technique to detect and classify the tumour part from medical brain images. In recent times, a number of image segmentation and detections techniques have been proposed in the literature. But, the detection of brain tumour through the help of classification technique has received significant interest among the research community. By considering the above issue, here, we combine three different techniques such as, cuckoo search, neural network and fuzzy classifier to detect the tumour part effectively. Our proposed approach consists of four phases, such as, pre-processing, region segmentation, feature extraction and classification. In the pre-processing phase, the anisotropic filter is used for reducing the noise and in the segmentation process; K-means clustering technique is applied. For the feature extraction, the parameters such as contrast, energy and gain are extracted. In classification, a modified technique called Cuckoo-Neuro Fuzzy (CNF) algorithm is developed and applied to detection of tumour region. In the modified algorithm, cuckoo search algorithm is employed for training the neural network and the fuzzy rules are generated according to the weights of the training sets. Then, classification is done based on the fuzzy rules generated. Experimental results shows that the proposed technique achieved the accuracy of 79.49% but existing technique achieved only 76.92%.

Keywords: CNF, contrast, energy, entropy, K-Means, anisotropic filter, sensitivity, specificity, accuracy

61. PaperID 310516183: Permission Based Android Malware Detection System using Machine Learning Approach (pp. 465-470)

Mayuri Magdum, Computer Engineering, Modern Education Society's College of Engineering, Pune, India
Prof. Sharmila K. Wagh, Computer Engineering, Modern Education Society's College of Engineering, Pune, India

Abstract — Mobile computing has grown and developed in recent years with huge popularity. Gadgets like Smart phones, Tablets, etc have become trendy by the ease of use. Android is more famous platform and turned out to be the most important target of Malware developers in precedent years. The malware hazard for cellular telephones is evaluated to increment security and usefulness of smartphones. Hackers and malware program developers are benefitted by the limited capabilities and lack of standard security mechanism of Android. Nowadays smart phones are omnipresent, i.e. they fill numerous needs such as data storage, personal mobile communication, multimedia and entertainment etc. therefore, implementing secure mobile connections is challenging. As a result, it becomes essential to have some valuable and probabilistic detection along with preventive mechanisms. Many preventive tools are available in market but current trend for malware security is before installing the app user should be able to identify possible threats. Hence we propose permission based mobile malware detection system. It has 3 components in it 1) Client 2) Server 3) Signature Database. In the whole analysis process, Server plays important role and user is warned at the end of analysis process whether the requested app contains malware or not.

Keywords- Mobile, Android, Malware, Security, Machine Learning, Static Analysis.

62. PaperID 310516190: Analysis of Decision Making Factors for Automated Intrusion Response System (AIRS): A Review (pp. 471-478)

Dileep Kumar Singh, Praveen Kaushik
Dept of CSE, MANIT Bhopal, India

Abstract — Increasing amount of dependability on computer networks and internet services are also increasing intrusions. Intrusion Detection System (IDS) tools detect the intrusions and produce alerts. An automated Intrusion Response System (AIRS) is required to analyze the alert and trigger appropriate response to mitigate the intrusion without delay. In this paper, cost evaluation methods and response decision making capabilities of various AIRS models are analyzed. Various decision making factors that are involved in the response selection process are also identified and then categorized in response, attack and system level factors.

Index Terms—Intrusion Response System, AIRS, Response selection, Response factors, Response cost.

63. PaperID 310516192: SQL Injection Prevention using Query Dictionary Based Mechanism (pp. 479-485)

Adwan F. Yasin, Nael Zidan
Department of Computer Science, Arab American University, Jenin, Palestine.

Abstract — SQL Injection Attack (SQLIA) is a technique of code injection, used to attack data driven applications especially front end web applications, in which heinous SQL statements are inserted (injected) into an entry field, web URL, or web request for execution. "Query Dictionary Based Mechanism" which help detection of malicious SQL statements by storing a small pattern of each application query in an application on a unique document, file, or table with a small size, secure manner, and high performance. This mechanism plays an effective manner for detecting and preventing of SQL Injection Attack (SQLIA), without impact of application functions and performance on executing and retrieving data. In this paper we proposed a solution for detecting and preventing SQLIAs by using Query Dictionary Based Mechanism.

Index Terms—SQL Injection Attack, SQL Injection Attack Detection, SQL Injection Attack Prevention, Query Dictionary.

64. PaperID 310516195: An Optimized Approach toward Intrusion Detection Using Cluster-Like Behavior of Attacks (pp. 486-490)

Aliakbar Tajari Siahmarzkooh, Jaber Karimpour, Shahriar Lotfi
Faculty of Mathematical Sciences, Department of Computer Science, Iran, Tabriz, University of Tabriz

Abstract — Most of intrusion detection researches suffer from the following drawbacks: Dependencies between network nodes and cluster-like behavior of anomalies. Hence, this paper proposes a cluster-based approach in which the anomalies are detected using a new criterion related to the behavior of attacks. In addition, we provide a cluster-based data set which uses the flow-based data and graph properties to model the network traffic over time. The data set is built over the DARPA. Moreover, the anomalies are revealed by means of a criterion which is computed from internal and external weight of clusters. Finally, the proposed approach is evaluated and compared to other approaches. The evaluation results show the preference of our approach relative to other ones.

Keywords- Anomaly; DARPA data set; flow; graph clustering; intrusion detection

65. PaperID 310516197: A Comparative Study of Smoothing a Vehicle's Trajectory which is calculated by an Evolutionary Algorithm (pp. 491-496)

B.A. Buran is with the Department of Basic Sciences, Turkish Air Force Academy, Istanbul, Turkey.
S.H. Caglar is with the Department of Mathematics and Computer Science, Istanbul Kultur University, Istanbul, Turkey.
O.K. Sahingoz is with the Department of Computer Engineering, Turkish Air Force Academy, Istanbul, Turkey

Abstract — Determining a vehicle's trajectory is a complex and hard to solve type problem in the literature and it is identified as a NP-Hard optimization problem which is studied in different engineering disciplines such as computer, electrical and industrial engineering. It has been observed that such complex problems can be solved by using various approaches and lots of them are focused on the usage of Evolutionary Algorithms especially in case of a large number of controls points which are needed to be visited. Although these algorithms provide near optimal solutions, in the real world, vehicles are not able to follow this determined path (trajectory) without any deviation. Because vehicles are moving objects and each one moves with a certain speed. Therefore it is impossible for a vehicle to make a sharp turn after visiting control points. These vehicles need to make smoothed turns over these points. Therefore there will be a certain difference between the calculated path and the real path. It is needed to determine the real path by using necessary mathematical solutions for smoothing these paths. To ensure the motion continuity of vehicles, they need to follow paths determined according to a certain criterion. In this study, the most common smoothing methods which are used to ensure these continuities (Bezier, B-Spline and Dubins) have been compared and it is aimed to show the different approaches in an application area of path planning problems as a comparative study.

Keywords — Unmanned Aerial Vehicle, Path Planning Evolutionary Algorithm, Bezier Curves; B-Spline Curves, Dubins Path.

66. PaperID 310516200: Location-Based Routing Protocols GAF and its enhanced versions in Wireless Sensor Network a Survey (pp. 497-504)

Hanane AZNAOUI, Said RAGHAY & Layla AZIZ
Laboratory (LAMAI), Faculty of sciences and technology, University Cadi Ayyad Marrakech, Morocco

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

Abstract - Since the last two centuries, humanity has made scale steps in this attraction to innovation and technological progress. The emergence of global networks of computers corresponding to Wireless Sensor network WSN is one of those great steps that man could do. WSN is an advanced technology that occur in response to overcome user needs. It resolves many problem such as, controlling phenomena, monitoring places, and diagnostic. Nevertheless, this

advanced technology still incomplete in order to different constraints such as energy consumption, routing, aggregated data and security, also routing information represents a critical issue in it. For that, great researches designed. In this paper, we present a survey of GAF and their enhanced versions as Location-Based routing protocols in WSN, which allows reducing the consumed energy in the network and prolonging the network lifetime.

Keywords: WSN, routingprotocols, location-based, GAF.

67. PaperID 310516201: Comparison of RC2 and AES Using Windows Azure for Data Security in Cloud (pp. 505-509)

Muhammad Inaam ul haq, Department of Computer Science COMSATS, Institute of Information Technology, Sahiwal, Pakistan
Muhammad Umar, Department of Computer Science COMSATS, Institute of Information Technology, Sahiwal, Pakistan
Syed Muhammad Owais, Department of Computer Science COMSATS, Institute of Information Technology, Sahiwal, Pakistan

Abstract - Cryptography is a very useful tool to protect the properties of data like integrity, privacy, confidentiality in any environment. This paper explores some useful aspects of cryptography in cloud computing environment. There are different types of encryption algorithms used in order to ensure the data security. These algorithms are of different types like symmetric, a symmetric and hashing algorithms. The objective of this paper is performance analysis of selected set of algorithms on the basis of different parameters, so that the best out of all these options is chosen or combinations of some of them can be utilized to secure data in cloud computing environment. The algorithms included in this study are RC2 and AES. The parameters which are used for performance analysis are running time of the algorithm, data encryption capacity. These are the performance parameters which are calculated for every algorithm in cloud based environment i.e. windows azure simulator by utilizing visual studio IDE and profiler services by integrating windows azure SDK. The interpretation of these results are done by using various graphs which shows trend of a particular algorithms on basis of time of encryption and decryption.

Keywords: Cryptography, Cloud Security, RC2, AES, Windows Azure

68. PaperID 31011659: Optimized and Secure Authentication Proxy Mobile IPv6 (OS-PMIPv6) Scheme for Reducing Packet Loss (pp. 510-515)

Arun Kumar Tripathi, Department of Computer Science and Engineering, KIET Group of Institutions,Ghaziabad, India
R. Radhakrishnan, Department of Computer Science and Engineering, ABES Engineering College, Ghaziabad, India
J.S. Lather, Department of Electrical Engineering, National Institute of Technology, Kurukshetra, India

Abstract — Due to continuous evolution in hand handled mobile devices such as Smartphones, Laptops, tablets and Personal Digital Assistants (PDAs) have increases the volume of traffic on Internet radically. To provide seamless Internet services and perpetual mobility to these devices, Internet Engineering Task Force (IETF) has proposed various mobility management protocols such as MIPv6, HMIPv6, and PMIPv6. MIPv6 is a host-based mobility management protocol and suffers from handover latency, packet loss etc. Recently the IETF proposed network-based mobility management protocol, known as Proxy Mobile IPv6 (PMIPv6). PMIPv6 sufficiently reduces signaling overhead but still have long authentication latency during handover and packet loss issues. To resolve these issues, an optimized and secure authentication mechanism for handover management scheme for PMIPv6 networks is proposed in this paper. Due to less authentication delay, the proposed scheme reduces the setup time and as a result has low handover latency. Subsequently, decreases the amount of packet loss during handover. The proposed scheme provides higher security infrastructure than the basic PMIPv6 protocol and additionally reduces the handover latency to contemporary protocols. The performance and results are mathematically analyzed. Numerical results show that the proposed scheme

gives better performance than the existing MIPv6 in terms of signaling delay and provide higher security than PMIPv6 protocol.

69. PaperID 31051605: Design for ALL: Catering for Culturally Diverse Users (pp. 516-524)

Mahdi H. Miraz (1,2), Maaruf Ali (2), Peter S. Excell (1)
(1) Department of Computing, Glyndŵr University, Wrexham, UK
(2) Department of Computer Science & Software Engineering, University of Hail, KSA

Abstract — Due to mass global migration and increased usage of the Internet, it is now very important to address the cultural aspects of the usability problems of any Information and Communication Technology (ICT) products such as software, websites or applications (apps) whether to be used on PCs, Laptops, Smartphones, Tablets, Smart TVs or any other devices. To augment the "Design for All" concept, this research demonstrates the need to cater for culturally diverse users while designing user interfaces. This has been achieved, by investigating ICT products and conducting an extensive literature survey. The study concludes that it is very important to work on cross-cultural usability problems and bring these issues under focus.

Index Terms — Human Computer Interaction (HCI), Universal Usability, Cross-cultural Usability, User Interface (UI) Design, Design for All, Users' Behaviour.

70. PaperID 31051611: Urban Traffic Control with Pedestrian Handling (pp. 525-534)

Ayesha Shahid, Wilayat Ali Khan, Nazish Naheed, Salman Hafeez, Syeda Nuzhat Subi Naqvi, Aihab Khan
Department of Software Engineering, Foundation University Rawalpindi Campus

Abstract - Over the years road traffic flow has seen pedestrian crossing as a major issue in the society, particularly in urban areas where there is no control for pedestrian road crossing. In mixed traffic conditions pedestrian road crossing behavior is a serious hazard for pedestrians crossing uncontrolled bi-intersection localities. Due to increase in motor vehicle growth there is an increase in the regulation of motor vehicles only and the regulation of pedestrian is completely neglected in urban area. An increase the uncontrolled road crossing behavior of pedestrian is raises different safety and economic concerns. This paper employs computational modeling to regulate the traffic flow across a two way intersection. It is caters how pedestrians can cross a bi-intersection traffic signal without disrupting the traffic flow. Existing computational models that have been presented by other authors are discussed which gives more understanding how to control traffic flow for vehicles and pedestrians handling. This study deals three scenarios of real environment for control of traffic flow for pedestrians; with no turns, with turns and with turns. All scenarios provides proper notation for 'on states' and 'off states' of signal. Experimental result demonstrates that the proposed method achieved waiting time for vehicles 143.35 seconds and 200.23 seconds for pedestrians respectively. Furthermore, result shows the decrement of time and economical resources that are used in the daily commute.

Index Terms— Pedestrian, Bi-intersection, uncontrolled traffic, Computational Modeling, Traffic Control System

71. PaperID 31051625: New Image Encryption Technique Based on Wavelet / DCT Transforms Using Lorenz Chaotic Map (NIETWDL) (pp. 535-547)

(1) Abdullah M. Awad, (2) Rehab F. Hassan, (1) Ali M. Sagheer
(1) University of Anbar, Ramadi, Iraq
(2) University of Technology, Baghdad, Iraq

Abstract - In communication networks, the data encryption has been used to safe the security of information. There are different encryption techniques that can be used to protect the data from unauthorized third person to access. This paper deals with chaos image encryption environment to hide the secret information and make communication undetectable. In this paper integer wavelet transform (IWT) and discrete cosine transform are used for increasing

hiding pixel distribution. The work uses IWT and DCT as a decorrelation stage for adjacent pixels. The performance evaluation for the proposed algorithm has been done by measuring the application using a series of tests. The tests include histogram analysis and visual test, correlation analysis encryption quality, information entropy, randomness test, sensitivity analysis and differential analysis. The proposed cipher algorithm experimental results show satisfactory security and efficiency levels for image encryption.

Keywords: Chaotic Encryption; AES; RC4; Statistical Analysis

72. PaperID 31051626: Stability Analysis of Reliable Ensemble Classifiers (pp. 548-557)

Zeinab Khatoun Pourtaheri, Seyed Hamid Zahiri, Seyed Mohammad Razavi
Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran

Abstract - In this paper, Multi-Objective Inclined Planes Optimization (MOIPO) algorithm, as a novel multi-objective technique, is used to design ensemble classifiers with high reliability and high diversity. It is noteworthy that sometimes, the reliability in decision of a classifier is more important than its recognition rate. Security and military applications are obvious instances to show the importance of this measure. In addition to reliability, diversity, as a main issue in ensemble classifiers, is considered as objective function. So, designing heuristic ensemble classifiers with high reliability and also, high diversity has a special importance but the basic point is that the applied heuristic algorithm has a stochastic nature and hence, stability analysis of this system is necessary. In this research, statistical method is used to do stability analysis of designed ensemble classifier.

73. PaperID 31051628: Design an Adaptive Kalman Filter for INS/GPS based Navigation for a Vehicular System (pp. 558-567)

Mohammad Nasiri, Department of Electrical Engineering, Faculty of Engineering, University of Birjand, Iran
Seyed-Hamid Zahiri, Department of Electrical Engineering, Faculty of Engineering, University of Birjand, Iran
Ramezan Havangi, Department of Electrical Engineering, Faculty of Engineering, University of Birjand, Iran

Abstract — Kalman filter is a very effective approach for data fusion. But, the definition of process, measurement noises, and the matrices Q, R have a great impact on the filter performance. Research works show that adjustment of matrices Q, R during the prediction process is very useful to reduce the estimation errors. So, in this paper, we attempt to increase the accuracy of Kalman filter used in INS/GPS integration algorithm by estimating measurement covariance matrix, R, based on measurement data from GPS. Our objective is to show a performance enhancement of a conventional extended Kalman filter used in an INS/GPS integrated navigation system by adjusting adaptively measurement noise covariance matrix R. This adaptive adjustment is necessary. Because, environment conditions in many systems usually are not constant and change continually.

Index Terms— Integrated navigation, Extended Kalman filter, Adaptive Kalman filter

74. PaperID 31051642: Efficient Image Enhancement Using Image Mining and Hadoop MapReduce (pp. 568-575)

M. Anand, V. Mathivanan
AMET University, Kanathur - 603 112 Chennai, India

Abstract - Multimedia has become part of our day today life especially when it comes as images. Many studies have proved that images are the most efficient way of expressing our feelings rather than a page of paragraphs. An example we can state here is the smileys we use in our messages for expressing our thoughts. The ultimate rise of social websites like Google+, Twitter and Facebook, playing major role in the Internet World has proved it wright since these websites are rich in content and huge number of images shared. The revolutionary technology development in the mobile industry is also playing the major role in using such multimedia content. Since the images are being shared in different

ways, people start compressing the images to reduce the huge amount of memory space. This compression leads to data loss (pixel) in images which affects the quality of the images. Many solutions have been identified to solve the issues. One such system uses one dimensional approach in all four directions (Row, Column, Diagonal and Inverse Diagonal); the recovery process is performed by considering the edge pattern of the existing image adjacent to the damaged data (pixel). The system also uses the method of determining the weighted sum [1] of selected point functions. Many more techniques followed like enhancement performed using: Spatial and Time domain [1], Frequency Domain Techniques [1], Brightness Preserving Bi-Histogram Equalization (BBHE) [2].

Keywords: Image Enhancement, Data Loss, Recovery process

75. PaperID 31051646: An Efficient Image Encryption Technique by Using Cascaded Combined Permutation (pp. 576-588)

Muslim Mohsin Khudhair
Department of Computer Information Systems, College of Computer Science & Information Technology, University of Basrah, Basrah, IRAQ

Abstract - In this paper, a new simple encryption technique is proposed for gray scale image encryption. The current technique, Cascaded Combined Permutation (CCP), is a simple technique based on the primary well known 2-D permutation algorithms. The application at the permutations is performed on three steps: (1) one permutation algorithm is applied on the image; (2) the image that resulting from the first step is decomposed into four quarters. Pixels in each quarter image are then permuted with one of the permutation algorithms. The resulting encrypted quarters are combined as one image; (3) the encrypted image resulting from the second step is further encrypted by performing another permutation algorithm. Experimental results show efficient encryption that is simple in implementation and has high degree of security. It has several key points of strength such as the sequence in which the primary permutation algorithms are applied.

Keywords: Permutation, Image Encryption, Image Decryption, correlation.

76. PaperID 31051658: Component Based Face Recognition using Feature Matching through Hu Moment Invariants (pp. 589-604)

Sushil Kumar Paul, Mohammad Shorif Uddin
Department of Computer Science and Engineering, Jahangirnagar University, Dhaka-1342, Bangladesh
Saida Bouakaz, LIRIS Lab, University Claude Bernard Lyon1, 69622 Villeurbanne Cedex, France

Abstract — In this paper, a Face Recognition Algorithm using Hu moment invariants (HMIs) is described for identifying human faces based on the facial component-features (FCFs). Algorithm is adopted by Viola Jones detector which is applied the concept on the AdaBoost algorithm for detecting the face from a face database having diverse illuminations and expressions with complex background. Then only the face region is cropped and illumination correction is done using histogram equalization technique. Finally, face is converted into binary image by applying cumulative distribution function (CDF) with adaptive thresholding. Three types of statistical pattern matching tools such as Standard deviation of Hu moment invariants (StdDevHMI), absolute difference of probability of white pixels (AbsDiffPWP) and pixel brightness values (PBVs) through L2 norms are determined using five facial components such as two eyes, nose, mouth and whole face for both binary and gray level images, respectively. Lastly, face recognition is carried out by taking these statistical pattern matching tools with logical and conditional operators along with appropriate threshold values. Experimental studies are performed on the BioID database and algorithm shows a better result as compare to the existing popular methods.

Keywords -- Cumulative distribution function, adaptive thresholding, probability of white pixels, facial component-features, shape matching, Hu moment invariants, pixel brightness values.

77. PaperID 31051664: A Robust and Efficient Optical Flow Analysis Based Vehicle Detection and Tracking System for Intelligent Transport System (pp. 605-613)

S Sri Harsha, Department of IT, VR Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India
K. R. Anne, Director Academics, Veltech University, Chennai, Tamilnadu, India

Abstract - In this paper, an enhanced optical flow analysis based moving vehicle detection and tracking system has been developed. A novel multidirectional brightness-intensity constraints (MBIGC) estimation and fusion based optical flow analysis (MDFOA) technique has been proposed that performs simultaneous pixel's intensity and velocity estimation in a moving frame for detecting and tracking the moving vehicle. The conventional Lucas Kanade and Horn Schunck optical flow analysis algorithms have been enhanced by incorporating a multidirectional BIGC estimation, which has been further enriched with a non-linear adaptive median filter based denoising. Such novelties have significantly enhanced the video segmentation and detection. A vector magnitude threshold based MDOFA algorithm has been developed for motion vector retrieval that eventually enables swift and precise moving vehicle segmentation from the background frame. A heuristic filtering based blog analysis has been applied for vehicle tracking. The MATLAB based simulation reveals that MDFOA-HS outperforms LK in terms of execution time and detection accuracy. In addition, the accurate traffic density estimation affirms robustness of the proposed system to be used in intelligent transport system.

Keywords: Multidirectional brightness-intensity constraint Optical flow analysis, intelligent transport system, Lucas Kanade, Horn Schunck.

78. PaperID 31051681: Area Efficient Digital Logic Circuits based on 5-input Majority Gate Using QCA (pp. 614-623)

D. Ajitha, JNTUA, Ananthapuramu, India
A. Harika, SITAMS, Chittoor, India

Abstract - Quantum-dot Cellular Automata (QCA) is one of the most significant technology among the Nano devices for computing at the Nanoscale. The key logic elements in QCA are majority gate and inverter. The majority gates are 3-input majority gate and 5-input majority gate. In earlier designs all the digital logic circuits are implemented using 3-input majority gate based on 2:1 multiplexer. The limitations of the 3-input majority gate are it requires the number of cells for constructing large architectures involves high complexity, connectivity is difficult, laborious and low reliability. Hence, the design of digital circuits in this paper is implemented with 5-input majority gate based 2:1 multiplexer. The 5-input majority gate reduces cell counts, the number of clocks required and area compared to existing designs. The proposed designs such as XOR gate, XNOR gate, D-latch, D flip-flop, T-latch, and T flip-flop have significant improvements regarding the number of gates, cell count, and delay. The proposed circuits are simulated with QCADesigner and results were included to verify the functionality.

Keywords: Quantum-dot Cellular Automata (QCA), Five-input Majority gate, Multiplexer, Logic gates, Sequential logic.

79. PaperID 31051689: Human Emotion Recognition and Prediction Using Socialism Media (pp. 624-633)

Dr. Tanvir Ahmed, Tamoor Khan
Lahore Leads University, Pakistan, Department of Computer Science

Abstract - Humans are unpredictable; there is no exact way or definition of emotion prediction. Detection of human emotion is difficult because when we want to observe people's behavior then they behave in normal way or better than abnormal behavior. May be another way where people want to collaborate with others to share their emotions, their daily basis problems, where they feel easy to share their expression without any fear. Maximum people are not agreeing to share their emotion due to shame and fear. We need a platform where people can share their actual problem (which they are internally facing) and release their frustration. Many people want solution without sharing of their

problems to anyone. In order to solve this problem, social media is a best way where people can share their emotional behavior without any fear and we can detect their emotion as silent observer through social media. In this paper we will analyze their posted data on social media and we have provided the suggestion to solve their problems; also we detected the emotion of people through social media. We collected data from social website (Twitter .etc.) where people have shared their thoughts or feelings. Meanwhile, we designed an algorithm which takes data from that social website and on the basis of that data; application provides the result as previous emotional state of a person. A systematic approach was used to detect the emotion of people through social media data. This is a better way where a person wants to collaborate with other to share his emotions, his daily basis problems and he feels easy to share his expression without getting panic. This Emotional based approach described things in a new way, where all predictions can be measured according to the subject environment and application can provide better results in decision making. This approach has used the data from social portals like Twitter etc. where peoples are posting their data in form of emotions. Prediction and recognition of emotions is a better way to analyze the emotion of people as silent observers.

Keywords — Emotion, Silent Observer, Parts of Speech (POS), Social Media(SM), Adjective

80. PaperID 31051690: Using Adaptive Filters for Object Tracking and Improving the Method Using Metaheuristic (pp. 634-640)

Farzad Khadem Mohtaram, Majid Mohammadi
Department of Computer, Buin-Zahra Branch, Islamic Azad University, Buin-Zahra, Iran

Abstract - The video detection based on the image sequence of the area of interest has attracted considerable attention. Particles filtration is one of the most development algorithms particularly in restoration of probability density function of goal state. Accordingly, the main objective of present study is utilization of adaptive algorithm for detection of inflexible objects. The simulation method was applied and data analysis is done by MATLAB software. The results represent that, filtration of the suggested particle achieved better performance than filtration of the standard particle in terms of prediction error of status, detection of video error, and the number of significant particles. It revealed that, the particle filtering enhanced the number of significant particles by IGA and, forced the collection of particles to better expression of actual status. This could enhance the accuracy of status prediction and reduced the error.

Keywords: adaptive algorithm, inflexible, objects detection, particle filtration

81. PaperID 31051695: Agile Practicing and Outsourcing (pp. 641-648)

Muhammad Sarfraz (1), Maria Ramzan (2), Akhter Rasheed (3), Fateh Ali (1)
(1) Department of Mathematics and Statistics, Riphah International University Islamabad, Pakistan
(2) Govt. Post Graduate College for Women Satellite Town Rawalpindi, Pakistan
(3) Department of Mathematics, COMSATS Institute of Information Technology Abbottabad, Pakistan

Abstract - The software industry can be widely seen as a key driver for business improvement. Outsourcing of software development tasks has become a major issue for large software enterprises. Software outsourcing has been progressively increasing. However significant outsourcing failure rates have also been reported. Therefore, outsourcing occurred by the wrong decision can cause major technological and economic setbacks. The objective of this research is to develop a model for outsourcing in order to improve outsourcing process and to help out the organizations to overcome barriers (communication, coordination & quality) that may have a negative impact on software outsourcing as well as to improve their success rate. Literature is consulted to highlight various issues of outsourcing. A case study is conducted to validate the effectiveness of our proposed model. The purposed model contains different practices of agile which provide an effective way to improve coordination, quality assurance and reduces communication gaps in outsourcing.

Index Terms- Agile, Outsourcing.

82. PaperID 310516101: Model Driven Architecture for Secure Software Development Life Cycle (pp. 649-661)

Muhammad Asad, Shafique Ahmed
National University of Sciences and Technology, Islamabad

Abstract – Secure Software Development is an important issues for the software industry for couple of years as security issues in the software development life cycle are not easy to handle. Success of a software deeply depends on the fact that it is not easily vulnerable to security threats and breaches. Many organizations have made security guidelines to cope with these challenges to bring them in an organized and secure way. Besides so much advancements in the field, securing the software from vulnerabilities in not achieved in all modules of software development life cycle. The guidelines and methods designed for the secure software development have put a lot contributions but they are so verbose that these measures are nearly not implementable. In this paper a model is proposed for secure software development life cycle in model driven architecture level (MDA-SDLC). In the proposed model, modeling methods and approaches are used to ensure the advances in secure model driven architecture with simplified integrity of security modules in security critical software's development lifecycles.

Keywords — Model Driven Architecture, Security, SDLC, UML,

83. PaperID 310516108: Social Relation Based Recommendation System For Information Overloaded Social Networks (pp. 662-671)

S. Uma Shankari, Bhararthyiar University, Coimbatore
Dr. M. Chidambaram, Rajah serofiji College, Thanjavur

Abstract - Social persuade plays vital part in the product marketing. Though, it's seldom been regarded in traditional Recommender systems (RS). This paper provides new paradigm RS which can exploit data in the social networks, with general approval of items, user preferences, and persuade from the social friends. The probabilistic representation is improved to build personalized recommendations like data. In world e-marketing, new commerce representations are normally introduced, new tendency started to materialize. Latest trend is the social networking websites, several of which concerned not only huge number of visitors and users, however online advertise company to put their ads on sites. This paper discovers online social networking like new e-marketing trend. We first inspect online social network like new web-based services, also evaluate social networks by other delegate web-based service. We extort information from real online social network, also our investigation of this huge dataset expose that friends contain tendency to choose similar items and provide similar ratings. The experimental outcome on the dataset illustrates that proposed scheme not only progress prediction accuracy RS but gives solution cold-start and data sparsity problems intrinsic in the collaborative filtering. Moreover, we recommend improving system performance by concern social networks semantic filtering, and authenticate its improvement through class project research. In this research we reveal how related friends may be choose for deduction based on the semantics friend relations and finer-grained customer ratings. Such technologies may be organized by mainly content providers.

Keywords: Recommender systems, collaborative filtering, social network

84. PaperID 310516109: Software Reengineering - A Frame of Reference (pp. 672-678)

A. Cathreen Graciamary, Bhararthyiar University, Coimbatore
Dr. Chidambaram, Rajah Serofiji College, Thanjavur

Abstract — Now day development of software is describe by immediate process. Old systems have to take on the recent technologies; It can be achieved by changing or finding the features, I.e, Reengineering. Our proposed paper clarifies about the reengineering process of software. It also explains the efficient and better process in reengineering. There are two type common reengineering objectives. Improved feature: the existing software system will be of minimum quality, because of more changing during the time course. The main objective of reengineering is to increase

software quality and to provide present working documentation. A higher quality degree is needed to enhance reliability, to minimize the maintenance cost, to develop maintainability, and to make for functional improvement.

Keyword- Software Reengineering, Reverse Engineering, Enhanced Reengineering, SVM classification, Software component.

85. PaperID 310516137: Analyzing Virtualization based Energy Efficiency Techniques in Cloud Data Centers (pp. 679-686)

Beenish Gul, Fiaz Gul Khan, Iftikhar Ahmed
Department of Computer Science, COMSATS Institute of Information Technology, Abbottabad, Pakistan

Abstract - Cloud computing provides IT services to users worldwide, Data centers in Clouds consume large amount of Energy leading to highly effective costs. Therefore green energy computing is solution for decreasing operational costs. This survey presents efficient resource allocation and Scheduling algorithm/Techniques analyzed on different network parameters without compromising network performance and SLA constraints. Results are analyzed on different measures, providing a significant cost saving and improvement in Energy Efficiency.

Keywords: Data Centers, Virtualization, Consolidation, Virtual Machines, SLA

86. PaperID 310516145: Image Share Pane Tool: Image Sending Approach to Mobile via Bluetooth Device (pp. 687-690)

Farhan Ali Surahio, Awais Khan Jumani, Javed Ahmed Mahar, Hidayatullah Shaikh
Department of Computer Science, Shah Abdul Latif University, Khairpur Mirs, Sindh, Pakistan

Abstract — Nowadays, Microsoft Word is commonly used in various areas including industries and academia. Microsoft word has introduced great user friendly features, for instance, Screenshot and Screen Clipping, Smart lookup, Tell Me and others. Among them, Layout option button has given us to set objects with line in text. Furthermore, Different types of panes have provided for various tasks. Microsoft Word has given us a facility to greet with thumbnail image of every window you have opened at the moment. Many users while working on document need to insert or capturing images with Screenshot and Screen Clipping, they want to share inserted images to mobile via Bluetooth But, Users are disappointed because there is no any tool provided to accomplish that task and user takes a long procedure to apply for sharing images to mobile through the Bluetooth. This paper provides an application which helps users to send an inserted image via Bluetooth while working on Microsoft word and they do not to switch any window. By adding it into existing Microsoft Word it will helpful for people living across the world.

Keywords- Screen Clipping; Layout Option; Share Option Button; Share Image Pane; Image capture format type

87. PaperID 310516154: An Optimal Approach for Securing the Data in Cloud Storage using Block Division and Predicate Encryption (pp. 691-696)

P. Vijaya Bharati, Department of Computer Science & Engineering, Vignan's Institute of Engineering for Women Kapujaggarajupet, Visakhapatnam
Prof. (Dr.) T. Sita Mahalakshmi, Department of Computer Science & Engineering, Gitam college of Technology Gitam University, Visakhapatnam

Abstract — The "pay-as-you-go" cloud computing model is an efficient alternative to store the data at a cheaper cost. Ensuring data security in cloud computing platforms is critical and has become one of the most significant concerns in the emerging field of cloud computing. The location of the servers where the data is stored and being accessed are not known to the end user. There are many numbers of different security models and algorithms which are applied to secure the data stored in the cloud. While these techniques are very nice, we cannot really always tell that they are

"unhackable". Given enough time, brains and tools any technique might be breakable because the techniques are not fine grained. The existing algorithms have their own flaws and so in this paper we proposed a method that is been improved in such a way that the data stored on the cloud is secured. The proposed method initially uses a lossless block division which divides the data into blocks and then division is applied storing the remainder and the group to which it belongs to separately and later we apply predicate encryption scheme on the data to be stored (remainder data) in which the keys correspond to predicates and cipher texts are associated with attributes. The public key PK with an attribute 'x' is used to encrypt the text and the secret key SKf corresponding to predicate f can be used to decrypt a cipher text with attribute 'x' if and only if f(x)=1.

Keywords: Block Division, Predicate Encryption, Predicates, Attributes, Secret Key

88. PaperID 310516164: A Collaboration between Two Readers for Clustering and Identification in RFID systems (pp. 697-706)

OUAKASSE Fathia, Applied Mathematics and Computer Science Laboratory (LAMAI) Cadi Ayyad University Marrekesh, Morocco.
RAKRAK Said, Applied Mathematics and Computer Science Laboratory (LAMAI) Cadi Ayyad University Marrekesh, Morocco.

Abstract - Radio Frequency Identification RFID is one of the most important technologies used in the internet of things. It is increasingly used in various applications because of their high quality as well as their low costs; however the avoidance of collision of tags during the identification process represents a great challenge, especially when the number of tags is too large. In this paper we propose a new mechanism, based on Progressive Scanning Algorithm, to group tags in the interrogation zone of a reader. The proposed mechanism consists in the deployment of two readers having the same interrogation zone. Simulated results show that the proposed mechanism can appropriately achieve higher performance compared to other existing algorithms in terms of the number of time slots allowing identifying tags and effectively in terms of total time required to do this.

89. PaperID 310516177: Web Page Classification based on Context's Semantic Correlation (pp. 707-713)

Mehrdad Ranjbar-Khadivi (1), Mohammad-Reza Feizi-Derakhshi (2)
(1) Department of Computer, East Azerbaijan Science and Research Branch, Islamic Azad University, Tabriz, Iran
(1) Department of Computer Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran.
(2) Department of Computer Engineering, University of Tabriz, Tabriz, Iran.

Abstract - Automatic web pages' classification is one way to deal with the increasing range of the World Wide Web. Considering that most of the content of web pages is text, so classification based on text is seems to be an efficient solution. The methods used for text classification are usually based on the key words. But if illusive keywords appear within the web page, then the class of the webpage will not be properly diagnosed. Therefore, rather than paying attention to the words, it is needed to be given to content and words meaning. In this paper, a method based on content semantic correlation has been proposed. A text consists of paragraphs, sentences and words. In this study at first text is divided into its components and stop words is removed. Then, in order to forms the basis of the words, it will be needed to find the root of the words. The Hypernyms Tree of words can be extracted by using FARSNET. By using this method not only is the meaning of the terms considered but also there is no need to clarify the words. After extracting the Hypernyms Tree for all keywords, text feature vector is created. Then the similarity of the text to each of the available categories measured. Finally, KNN classification algorithm is used to recognize the right class of the webpage. The results show that by using this method, classification accuracy is increased by 0.17 in compared with other methods.

90. PaperID 310516178: Relevance Feedback in XML Retrieval Based on Classification of Elements (pp. 714-734)

Inès KAMOUN FOURATI, Mohamed TMAR, Abdelmajid BEN HAMADOU
Multimedia, InfoRmation systems and Advanced Computing Laboratory, SFAX, TUNISIA

Abstract - Unlike classical information retrieval systems, the systems that treat structured documents include the structural dimension through the document and query comparison. Thus, the relevant results are all elements that match the user needs rather than the entire document. In such a case, the document and query structure should be taken into account in the retrieval process as well as during the reformulation. Query reformulation should also include the structural dimension. In this paper, we propose an approach of query reformulation based on structural relevance feedback. We start from the original query and the fragments judged as relevant by the user. The analysis of the structure of document fragments and textual content of elements enables identify elements that match the user query and rebuild it during the relevance feedback step. The main goal of this paper is to show the impact query reformulation based on an analysis of the structure and content of each relevant element retrieved by an initial search process. Some experiments have been undertaken into a dataset provided by INEX to show the effectiveness of our proposals.

Keywords: Information retrieval; XML document; relevance feedback; Line of descent matrix; Classification.

91. PaperID 310516181: An Access Fairness Resource Provisioning of Services for Energy Efficiency in Wireless Cellular Ad-hoc Network (pp. 735-747)

Sridhara S. B., Assistant professor Department of Electronic and Communication, Rajiv Gandhi Institute of Technology, VTU, Bangalore, India
Dr. Ramesh B., Professor Department of Computer Science and Engineering, Malnad College of Engineering, VTU, Hassan, India

Abstract - The recent growth and development of smart phone technology have resulted in the growth of production of low cost smart phone devices. Due to the availability of low costs smart devices have resulted in increasing in the number of application and its user. The users in cellular network are mobile in nature and varied application services is been used such as FTP (File Transfer Protocol), VoIP (Voice over Internet Protocol), Multimedia services etc…which requires different data rate for each services. To assure a QoS (Quality of Services) for this kind of user application dynamic requirement and is a challenge that exists in existing wireless cellular adhoc network that need to be addressed. To achieve an efficient QoS & D2D (Device to Device) architecture is required. Many existing work based on D2D on cellular network have been proposed in recent times but they are not efficient in term of access fairness for varied traffic classes and it induces high cost of deployment since it require new infrastructure. To overcome this here the author adopts a cost effective D2D multicast communication based on pre-processed cellular infrastructure graph and admission control strategy for selectivity of services of varied traffic size in order achieve an efficient access fairness that reduces the packet drop rate and improves the overall packet delivery ratio of the network. The simulation outcomes show that the proposed model reduces the packet drop rate and improves the packet delivery ratio of the cellular ad-hoc network.

Keyword: Admission control, cellular network, graph pre-processing, d2d, routing.

92. PaperID 310516188: Decision Supporting Technique and Conventional Approaches – A Review (pp. 748-769)

Dr. S. Manju, Department of Computer Science, Sri Ramakrishna College of Arts and Science for Women, Coimbatore

Abstract - Brainstorming is a technique for generating a large number of ideas for creative problem solving. The generation of new ideas, especially high quality creative ideas is important for a problem. It is a popular method of group interaction in both educational and business sectors. Brainstorming engenders synergy i.e., an idea from one participant can trigger a new idea in another participant. Brainstorming must been recognized as an effective group decision supporting approach. This paper discusses about some of the variations of Brainstorming techniques and

previous approaches carried out to improve the quantity and quality of ideas, significance of creative thinking, target to increase productivity, requirement of group brainstorming and effectiveness of E-Brainstorming.

Keywords: Brainstorming, Decision Support System, Creativity, Management Information System.

93. PaperID 310516191: A Neural Network Model for Predicting Insulin Dosage for Diabetic Patients (pp. 770-777)

Bilal M. Zahran, Department of Computer Engineering, Faculty of Engineering Technology, Al-Balqaa Applied University, Jordan

Abstract - Diabetes Mellitus is a chronic metabolic disorder. Normally, with a proper adjusting of blood glucose levels (BGLs), diabetic patients could live a normal life without the risk of having serious complications that normally developed in the long run. However, blood glucose levels of most diabetic patients are not well controlled for many reasons. Although the traditional prevention techniques such as eating healthy food and conducting physical exercise are important for the diabetic patients to control their BGLs, however taking the proper amount of insulin dosage has the crucial rule in the treatment process. In this paper we have proposed a model based on artificial neural network (ANN) to predict the proper amount of insulin needed for the diabetic patient. The proposed model was trained and tested using several patients' data containing many factors such as weight, fast blood sugar and gender. The proposed model showed good results in predicting the appropriate amount of insulin dosage.

Keywords: Diabetes, Artificial Neural Network (ANN), Blood Glucose Levels (BGLs)

94. PaperID 310516193: An Optimum Dynamic Time Slicing Scheduling Algorithm Using Round Robin Approach (pp. 778-798)

Mohammad Salman Hafeez & Farhan Rasheed
Department of Software Engineering, Foundation University Islamabad, Pakistan

Abstract - Process Management is one of the primary tasks achieved by the Operating Systems. The system's performance sententiously depends upon CPU scheduling algorithms. Round Robin, contemplated as the most extensively endorsed CPU scheduling algorithm, is an optimal solution for the timeshared systems. In timeshared systems, selection of the time quantum plays a pivotal role in performance of CPU. In Round Robin, the static nature of the time quantum emerges some problems directly related to the quantum size which decreases the performance of CPU. In this paper, selection of time quantum is reviewed and a new algorithm for CPU scheduling, Optimum Dynamic Time Slicing Using Round Robin (ODTSRR) is proposed for timeshared systems. The proposed algorithm is based upon dynamic time quantum. Round Robin algorithm is redressed in this paper, ODTSRR also contains the advantages of RR (Round Robin) CPU scheduling algorithm have less chances of starvation. Performance of proposed algorithm is compared with RR and other shades of RR and the results revealed that the proposed algorithm is better in response time & waiting time, context switch rates, turnaround time and throughput hence resulting in optimized CPU performance.

Keywords: Operating System, Scheduling, Round Robin CPU scheduling algorithm, Time Quantum, Context switching, Response time,, Turnaround time, Waiting time, fairness.

95. PaperID 310516194: Profile Screening and Recommending using Natural Language Processing (NLP) and Leverage Hadoop Framework for Bigdata (pp. 799-811)

Mrs. D.N.V.S.L.S. Indira, Dr. R. Kiran Kumar
Dept. of Computer Science, Krishna University, Machilipatnam,

Abstract - Recommendation has been a major area that any recruiter would look for on a given job description. Increase in digital communication has made things easy to upload resumes and make it available for recruiters; on the other hand increase in technologies would make any recruiter difficult to scan it manually. Here we introduce an application which processes text data, understands sentence behavior unlike conventional keyword search applications and gives out required resume as per job description provided to application. This application makes use of Natural Language Processing (NLP) which helps in data training and feature extraction of the text data. Using NLP methods, semi structured text data is converted to structured format with required extracted features. To make this application scalable to any size of data we propose this implementation on Hadoop framework, which can handle any number of resumes or even more than petabytes of data, termed as bigdata.

Keywords: BigData, Attribute Tagger, NLP Methods, Named Entity Recognition (NER), Map-Reduce, Hadoop, HBase, Hive

96. PaperID 310516199: Real Time Variable Voltage Scaling to Design Energy Efficient Systems (pp. 812-820)

Ankita Soni & Praveen Kaushik
Department of Computer science and Engineering, Maulana Azad National Institute of Technology, Bhopal, India

Abstract - With the immense increase in the processing power over the past few decades, battery life has proved to be a crucial resource. Since energy varies quadratically with voltage in the CMOS based processors, Dynamic Voltage Scaling (DVS) offers a solution to conserve the battery power by lowering the supply voltage. However, reducing the voltage increases the execution time and therefore, real time scheduling has to be combined with DVS so as to provide the deadline guarantee. This paper presents an algorithm, Recurring Variable Voltage Scheduling(RVVS) to extend the battery life using a combination of variable voltage and a real time scheduling algorithm (Earliest Deadline First). The paper also mathematically proves that if two voltage levels are used such that one is twice the other, up to 50% energy can be saved. Mathematical proof of delay increment due to voltage reduction has also been presented. RVVS has been optimized in order to reduce the overall energy dissipated by switching by introducing a factor 'n' that denotes the number of time units after which the voltage switch can occur. RVVS has been applied to task sets having different number of tasks providing an average energy saving of 27%. This significant amount of energy saving helps extending the battery life to a remarkable extent and proves the worth of RVVS in the field of real time DVS.

Keywords: Dynamic Voltage Scaling; Earliest Deadline First; Real time scheduling; Voltage switching; Energy efficiency; Variable voltage

97. PaperID 310516202: Design and Detection of Network Covert Channel - An Overview (pp. 821-828)

R. Rajamenakshi, Department of Computer Science, Avinashilingam Deemed University, Coimbatore, India
Dr. G. Padmavathi, Department of Computer Science, Avinashilingam Deemed University, Coimbatore, India

Abstract - Sensitive information leakage is increasing due to wide spread use of internet and technology. The attackers find new ways to exfiltrate data that pose threat to data security and privacy. Here our focus is on the covert information leakage over the network that exploits the various network protocols and their behavior. Information leak over covert channels exploit a variety of protocols of network protocols including Wireless, mobile and virtualized cloud platforms etc. Current network security solutions like IDS, IPS, firewalls etc. are not designed to handle these type of attacks. These type of attacks are dynamic in nature and mimics the legitimate traffic behavior, there by posing a challenge to detect and prevent. This article presents comprehensive review of the network covert channel, design, detection and mitigation. We have reviewed the classification of covert channels based on the attacks.

98. PaperID 31051678: Generalized Intuitionistic Fuzzy Interior Ideals of Semigroups (pp. 829-836)

Muhammad Sajjad Ali Khan, Muhammad Shakeel, Khaista Rahman
Department of Mathematics, Hazara University, Mansehra, Pakistan

Saleem Abdullah, Department of Mathematics, Abdul Wali Khan University, Mardan, KPK, Pakistan

Abstract — In this paper we introduce and study a new sort of intuitionistic fuzzy interior -hyperideals of a -semihypergroup, called (,)-intuitionistic fuzzy interior -hyperideals by using the combined notions of belongingness and quasicoincidence of intuitionistic fuzzy points and intuitionistic fuzzy sets and some interesting properties are investigated. We show that an IFS A = ⟨A, A⟩ is an (∈, ∈∨q)-intuitionistic fuzzy interior -hyperideal of H if and only if U(t, s) ={x ∈ H: x(t, s) ∈A} for all t ∈ (0,0.5] and s ∈ [0.5, 1) is interior Γ -hyperideal of H. Moreover, we show that an IFS A = ⟨A, A⟩ is an (∈, ∈∨q)- intuitionistic fuzzy interior -hyperideal of H if and only if [A](t, s) ={x ∈ H: x(t, s) ∈ ∨qA}for all t ∈ (0, 1] and s ∈ [0, 1) is an interior -hyperideal of H. These showed that (∈, ∈∨q)-intuitionistic fuzzy interior -hyperideals of H are generalization of existence of intuitionistic fuzzy interior Γ-hyperideal of H.

Keywords: Semigroup, Intuitionistic fuzzy point; Intuitionistic fuzzy sets; (,)-Intuitionistic fuzzy interior ideal.

99. PaperID 310516138: Pythagorean Fuzzy Hybrid Geometric Aggeration Operator and Their Applications to Multiple Attribute Decision Making (pp. 837-854)

Khaista Rahman, Muhammad Sajjad Ali Khan, Muhammad Shakeel
Department of Mathematics, Hazara University, Mansehra, KPK, Pakistan
Saleem Abdullah, Department of Mathematics, Abdul Wali Khan University, Mardan, KPK, Pakistan

Abstract: There are many aggregation operators and its applications have been developed up to date, but in this paper, we develop the Pythagorean fuzzy hybrid geometric (PFHG) operator, and also study some properties, such as monotonicity, idempotency, and boundedness of the proposed operator. Pythagorean fuzzy hybrid geometric operator is the generalization of the Pythagorean fuzzy weighted geometric (PFWG) operator and the Pythagorean fuzzy ordered weighted geometric (PFOWG) operator. Finally, we apply the Pythagorean fuzzy hybrid geometric (PFHG) operator to deal with multiple attribute decision making (MADM) problems under Pythagorean fuzzy information. Using Pythagorean fuzzy hybrid geometric aggregation operator, we also develop an algorithm for multiple attribute decision making (MADM) problems. Lastly we construct an example for multiple attribute decision making MADM problems.

Key words: Pythagorean fuzzy sets, Pythagorean fuzzy hybrid geometric PFHG operator. Decision making problems.

100. PaperID 310516143: Cultural Factors Affecting ICT Acceptance Case Study: Industries Located in Science and Technology Park, Tehran (pp. 855-865)

*Mina Babazadeh Farokhran (*1), Behrouz Eskandarpour (2), Hossein Eskandarpour (3), Rogaye Rezaee Giglo (1),*
(1) Young Researchers and Elite Club, Germi Branch, Islamic Azad University, Germi, Iran
(2) Department Of Management Payamenoor University (Pnu) Iran,
(3) Zanjan Islamic Azad University, Zanjan, Iran

Abstract - Application of new technologies is considered as a key factor for the development of companies in recent years. This puts emphasis on the importance of reviewing factors influencing the acceptance of information technology culture. This study has been done aiming to identify factors influencing the information technology acceptance in companies located in the Tehran science and technology park. 80 companies from industries based in science and technology parks in Tehran were selected of these, 72 questionnaires have been evaluated and Cronbach's alpha was used to measure the reliability and validity of measurement tools. The reliability coefficient of the questionnaire is 0.86, which indicates high reliability of the applied questionnaire and content validity was confirmed by instructors. The research data is analyzed by SPSS which uses the correlation analysis along with significance levels and in the following, t and f tests have been used to study the research additional hypotheses. The results of this study showed that the usefulness and ease of use and subjective norms affect the information technology acceptance through

behavior intent and using independent ttest, it was found that looking at research indicators is alike among men and women. Based on the f statistics, attitude to these indices among different education levels is different and the respondents' education has an impact on attitudes to these indicators.

Keywords: cultural factors, Information Technology, technology acceptance, TAM, UTA

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 6, June 2016

ESSPI : Exponential Smoothing Seasonal Planting Index, A New Algorithm For Prediction Rainfall

Kristoko D. Hartomo
Faculty of Information Technology
Satya Wacana Christian University
Salatiga, Indonesia

Subanar
Faculty of Mathematics and Natural Sciences
GadjahMada University
Yogyakarta, Indonesia

Edi Winarko
Faculty of Mathematics and Natural Sciences
GadjahMada University
Yogyakarta, Indonesia

Abstract—**Exponential smoothing algorithm is a prediction algorithm recommended by the Food and Agriculture Organization. The weakness of the this exponential smoothing prediction algorithm is low accuracy for the prediction of long-term and ineffective in determining the value of smoothing to minimize error. The proposed research is to build a model rainfall prediction using a new algorithm Seasonal Planting Index (ESSPI). By using the algorithm planting seasonal index, rainfall prediction model will generate higher accuracy. The results showed seasonal planting method is the dominant index (5 of 6 test size) have an average accuracy is better than the method of exponential smoothing. Index planting seasonal prediction accuracy of 95.73% better than the exponential smoothing α = 0.1 by 56.55%, and exponential smoothing of α = 55.53. Novelty of this research is new algorithms for classifying data based on seasonal planting index, a new algorithm for determining the smoothing (value), the new fitting algorithm using seasonal planting index, and new algorithms using seasonal rainfall prediction planting index for the determination of the growing season.**

Keywords—exponential; smoothing; algorithm; seasonal planting index; predictions; accuracy; rainfall; novelty

I. INTRODUCTION

The shift of the rainfall patterns affect agricultural resources and infrastructure that led to the shifting of the growing season, seasons, cropping patterns, and land degradation. Changes in the rainfall patterns have occurred since the last few decades in some parts of Indonesia, such as the shift in the beginning of the rainy season and changes in rainfall patterns [1][2]. In addition there is the tendency of change in the intensity of monthly precipitation with the diversity and the deviation of the higher as well as increased frequency of extreme climate events, especially rainfall, wind, and tidal flood. With the trend of shortening of the rainy season and increased rainfall in the southern island of Java and Bali in Indonesia resulted in changes in the beginning and duration of the growing season, thereby affecting the planting index (IP), the planting area, early planting and cropping patterns [1]. Pullback beginning of the rainy season for 30 days can reduce rice production in West Java and Central Java as much as 6.5% and on the island of Bali reached 11% of normal conditions [2].

Exponential smoothing algorithm is inconsistent short-term forecasting, as an example of the decline in production of agricultural cultivation in an area caused by the incidence of drought, but the model still describes the increase in production [4]. The prediction results with exponential smoothing algorithm would be accurate only for a short-term prediction (prediction 1-2 periods) [3]. Determining the value of smoothing (value) must be made in advance to minimize the error in the exponential smoothing algorithm, commonly used method is trial and error [5]. Rainfall prediction models used to determine cropping patterns currently not able to provide the results of long-term rainfall prediction is accurate and still use methods with high level of complexity [6][7][8]. Exponential smoothing is a method that shows weighting decreases exponentially with observed values of the past, newer values assigned weights are relatively greater than the value of observation is longer [5].

The solution proposed in this study have relevance to the data pattern of rainfall in Indonesia, which generally have three main characteristics, namely high rainfall (January - April), low rainfall (May - August), and moderate rainfall (September - December). Despite the change in the time of rainfall during the period of last 15 years but the cropping of rice in Indonesia following the three characteristics of the rainfall. Rainfall prediction model proposed with respect to patterns of rice planting time become the main study in modeling, it contributes to the novelty of the existing exponential smoothing algorithm.

II. RELATED WORKS

The purpose of the time series concept application in general is to understand the behavior of the future through the measurement attribute of data in a time series of the past by using a trend indicator, cyclic and seasonal [9][10][11][12][13]. There are three categories of exponential models, Simple Exponential Smoothing (SES), Double Exponential

Smoothing (DES) and Triple Exponential Smoothing (TES). SES method, are used for short - term predictions, assuming that the data rate is still fluctuating around a mean value without the formation of a tendency (trend) is fixed [5], SES equation is as follows :

$$F_{t+1} = a * x_t + (1 - a) * F_t \tag{1}$$

F_{t+1} is the prediction for time *t+1*, $x_t + (1 - a)$ is the actual value of the time series, and a is a constant smoothing, denoted as α with a value between 0 and 1 [11], DES is used to form a pattern of the data that tendency (trend). Trend is defined as the result of smoothing the estimated average growth - average in each period of the data defined as follows :

$$S_t = a * y_t + (1 - a) * (S_{t-1} + b_{t-1}) \tag{2}$$
$$b_t = \gamma * (S_t - S_{t-1} + (1 - \gamma) * (b_t - 1) \tag{3}$$
$$F_t + m = S_t + b_t + m \tag{4}$$

S_t is forecast in period t, $y_t + (1 - a)$ is the actual value of the time series, bt is the value trend in period t, a is the first parameter of smoothing between 0 and 1 for smoothing the value of observation, γ is the second parameter for smoothing trend, $F_t + m$ is the prediction results to m and m is the number of periods ahead to be predicted [11]. TES is used for processing the data that is seasonal, formulated :

$$b_t = g(S_t - S_{t-1}) + (1 - g)b_{t-1} \tag{5}$$
$$I = b\frac{tX}{tS + (1-b)t - L + m} \tag{6}$$
$$F_t + m = (S_t + b_t m)I_t - L + m \tag{7}$$

b_t notation is the trend value at a certain period, S_t is forecast in period t, $F_t + m$ is the prediction results to m, m is the number of periods ahead to be predicted, L is the length of the season (number of seasons per year), $F_t + m$ is the outcome prediction at time t and m season period, and I is the seasonal adjustment [11].

ARIMA methods used to predict rainfall in Dhaka using size RMSE test that produces low value standard error. The weakness of this method of ARIMA for forecasting rainfall is high complexity in the implementation of ARIMA methods become rainfall prediction model [14][16]. Time series analysis and prediction has become a major tool in a variety of applications in meteorological phenomena, such as rainfall, humidity, and temperature [15]. The development of predictive models to estimate the parameters weighted by the least squares method on exponential smoothing model of hybridizing with the neural network does not provide optimal results [8]. This is due to the squaring errors in model fit merger will shift the curve to a point in another, thereby reducing the accuracy of the prediction.

From a literature review rainfall prediction can be analyzed that the application of exponential smoothing method for rainfall prediction of short-term (one or two periods) has good accuracy with low error and satisfies the principle of parsimony (simplicity) to be applied. Weakness exponential smoothing method is a large error value (inaccurate) prediction for the medium and long term. The other drawback is inefficiency mechanism of trial and error to determine the value of smoothing (value) [5]. While the general weakness encountered in the application of ARIMA method for forecasting rainfall is high complexity in the implementation of ARIMA methods become a model rainfall prediction.

III. Proposed Method

Figure 1 shows the proposed model of rainfall prediction algorithm exponential smoothing seasonal planting index is divided into three stages : (a) the first stage preliminary processing research data consists of the preprocessing stage rainfall data to purge the data from the noise and grouping data based on seasonal planting index, (b) the second phase to create a model rainfall prediction in two steps, determine the value of smoothing based on seasonal planting index and perform rainfall prediction algorithms using ESSPI, and (c) the third stage of testing error size models test ME, RMSE, MPE, MAPE, MASE, and Euclidean Error. Seasonal planting index represents the number of species groups measured quantity or hose replacement time (months) at the time of planting. By using seasonal planting index, exponential smoothing models produce higher accuracy.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

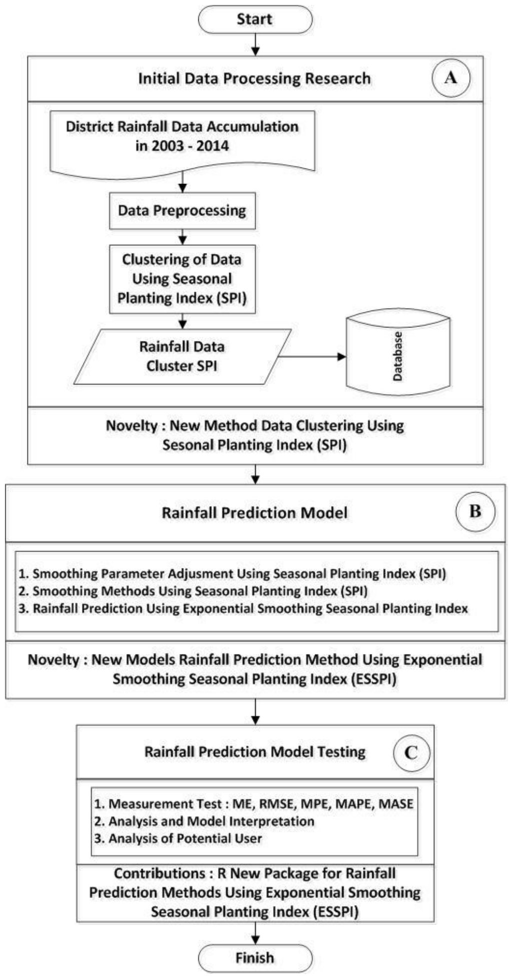


Fig. 1. Stage in making a prediction model precipitation seasonal exponential smoothing index planting

A. *Initial Data Processing Research*

The first stage of rainfall data preprocessing, the data used in this research come from Department of Forestry and Plantation Boyolali Regency which is the measurement results postal rainfall station in each district. The research used rainfall data districts (with a period of months) from 2003 to 2014. The use of monthly data for rainfall prediction is based on a literature study [16] [17] [19] [20] the more periods used in the pattern the data will be more easily analyzed to determine the correct prediction. Rainfall data sourced from the 19 districts in the study area of Boyolali Regency includes the district of Ampel, Andong, Banyudono, Boyolali, Cepogo, Juwangi, Karanggede, Kemusu, Klego, Mojosongo, Musuk, Ngemplak, Nogosari, Sambi, Palm, Selo, Simo, patio, and Wonosegoro.

Experiments performed on the data preprocessing district that has a missing value with the highest percentage reached 7.6% as Ngemplak districts while the average percentage of missing value 18 other districts in Boyolali was 4.5%. From the analysis of the data is needed rainfall preprocessing stage to obtain quality data, the quality of input data (noise-free), we get validity and accuracy output. The next process is data cleaning by doing replace missing value to clear the rainfall data of noise and distortion values. It starts with checking whether the input data has an empty/ missing value / tupple, if there is empty value of the attribute, empty value can be filled with new values using a linear interpolation method [21].

Data rainfall in Indonesia generally have three major characteristics such as high rainfall (around January - April), low rainfall (around May -August) and moderate rainfall (around September - December), it is seen in Figure 2. Although no change in the peak of rainfall in the last 12 years but the pattern of planting crops in Indonesia following 3 kinds of characteristics of the rainfall. Rainfall prediction model with respect to time of planting into the main study in the preparation of the model that contributes the newness to the exponential smoothing algorithm that already exists (exponential smoothing according brown, holt, and winter).

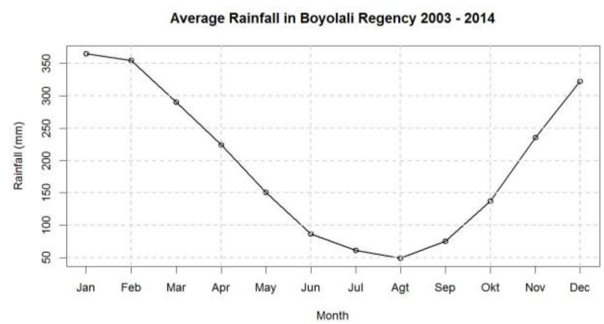


Fig. 2. Graph of the average monthly rainfall data Boyolali Regency Central Java Province, Indonesia

The Selection of exponential smoothing models using the base of Brown because it has a simple model [22]. Seasonal Planting Index (SPI) is the number of groups that are measured divided by a long planting (months) within a period of 12 months (one year). SPI is symbolized $I_{SP} = m/L$, where m is the number of groups that are measured (the number of the growing season) in one year and L is the length of planting (the time it takes from planting to harvesting). The measured data is rainfall, which are grouped into three types, such as: group precipitation in January - April called type 1, group precipitation in May - August called type 2, and the group precipitation in September - December called type 3. So the number of groups (planting) is $m = 3$, while the old plant (months) is $L = 4$ (corresponding to the period of paddy planting at the site of research) so determined $I_{SP} = 3/4$. Value I_{SP} is general with the provisions of rainfall data is grouped into m (according to the type of rainfall patterns / number of planting period in a year), while old crop planting (months) is L.

Data clustering based SPI (Seasonal Planting Index), based on the type of data that is grouped with SPI then each group is processed according to the equation 8 and equation 9, the rainfall data expressed in vector x.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$\vec{x} = \begin{cases} \vec{u} \\ \vec{v} \\ \vec{w} \end{cases}, \text{where } \vec{x} \in \mathcal{R}_{+}^{\mathrm{N}}, N = m * L \quad (8)$$

$$\vec{u} \in \mathcal{R}_{+}^{\mathrm{L}}, \vec{v} \in \mathcal{R}_{+}^{\mathrm{L}}, \vec{w} \in \mathcal{R}_{+}^{\mathrm{L}} \quad (9)$$

Where :
m = many clusters
N = lots of observational data (month)
L = many months in the first term of the rice planting
$\vec{u}, \vec{v}, \vec{w}$ = vectors in each different planting season

Algorithm 1 is used to classify the data according to the type of quantity of data in each cluster. For i = 1 if it is less than the amount of data as much as N then the value for cluster 1 on the index $j = i$ to $i + L$ (the number of months of planting) - 1 would be worth of data on the index - j. For cluster 2 will be worth of data on the index $j + L$ (the number of months of planting) * 1 and for cluster 3 will be worth of data on the index $j + L$ (the number of months of planting) * 2.

Algorithm 1 Grouping the data with Planting Seasonal Index (SPI)

Input: Data [N] = rainfall data area as many as N.
1. $L \leftarrow$ growing number of months (4 months)
2. $m \leftarrow$ many types of growing season (3 types)
3. $\text{Cluster} \leftarrow \text{Cluster1}\left[\frac{N}{m}\right], \text{Cluster2}\left[\frac{N}{m}\right], \text{Cluster3}\left[\frac{N}{m}\right]$
4. period $\leftarrow$ the number of months in the first year
5 $i \leftarrow 1$
6. **Do**
7. **For** $j = \mathrm{i}$ **to** $(i + L - 1)$ **do**
8. $\text{Cluster1}[j] \leftarrow \text{data}[j]$
9. $\text{Cluster2}[j] \leftarrow \text{data}[j + L * 1]$
10. $\text{Cluster3}[j] \leftarrow \text{data}[j + L * 2]$
11. **End For**
12. $i \leftarrow (i + \text{period})$
13. **While** $i \leq N$

Data Grouping is done as many as the number of months of planting in one period (line 1) and the process repeated clustering of data (line 7). With reference each iteration ends at one period it will be added the next period on the value of i (line 12) and re-check the condition of the value of i whether to re grouping the data or not (line 13).

B. Creating Prediction Model for Precipitation with Planting Seasonal Exponential Smoothing Algorithm Index

The second stage is to determine the value of smoothing (smoothing) based SPI, for the value of $\propto$ is $0 \leq \propto \leq 1$ [5] if $\propto = 0$, the function exponential smoothing will not be undefined. If $\propto = 1$ then $S_t = \propto \frac{x_t}{1_{t-L}}$ obviously this is not appropriate for the data to be predicted S_t only based on data x_t so that the data would have predicted the previous data, therefore, use the value smoothing $0 \leq \propto \leq 1$. The literature indicates the value of smoothing α obtained by trial and error; it is considered very ineffective to improve prediction accuracy quickly and accurately. In this research the proposed value of α formulated *ansatz* using the exponential function where $0 \leq \propto \leq 1$.

Smoothing method using SPI algorithm gives freshness to the definition of value as a weight smoothing parameter (equation 10) because in general they use trial error. By using the I_{SP} then the parameters α are symbolized as $\alpha_{I_{SP}}$ formulated :

$$\alpha_{I_{SP}} = 1 - \exp(-I_{SP}) \quad (10)$$

Exponential function (*exp*) is selected to determine the value smoothing (α) as the standard prediction methods used are exponential smoothing method. Smoothing value (α) must be between $0 \leq \propto \leq 1$ then have the power of negative. Based on analysis of the pattern of annual rainfall, the weather natural phenomena in the area of research, and the life cycle of rice (the time from planting to harvesting) within a period of 12 months (1 year), the determined value of the $I_{SP} = ¾$. Value 3 is the number of the planting season in the first year and a value of 4 is the number of months in one life cycle of the rice plant.

Value smoothing process results planting seasonal index will be used in the formulation of smoothing, as in the literature (formal call) that one type of smoothing method is Brown's linear exponential smoothing (LES) or Brown's double exponential smoothing. However, the revised standard formulation using weighting parameter based SPI wherein m in each year during the first year, the first test data satisfies the equation :

$$S_0' = x_0$$
$$S_0'' = x_0$$
$$S_t' = \alpha_{Isp} x_t + (1 - \alpha_{Isp}) S_{t-1}' \quad (11)$$
$$S_t'' = \alpha_{Isp} S_t' + (1 - \alpha_{Isp}) S_t' \quad (12)$$

For every $t = kL + 1$ then $S_t' = x_t$ for k = 1, ..., n and k = t + 1 where N is infinite, then the smoothing process is continued by defining :

$$b_t = \frac{\alpha_{Isp}}{1 - \alpha_{Isp}} (S_t' - S_t'') \quad (13)$$
$$a_t = S_t' + (S_t' - S_t'') = 2S_t' - S_t'' \quad (14)$$

The estimated value selected:

$$F_t = |a_t + I_{sp} b_t| \quad (15)$$

Where :
x_0 = actual data first
S_0' = smoothing or smoothing first single
S_0'' = double smoothing or smoothing the second
S_t' = single smoothing on the value of t
S_t'' = double smoothing on the value of t
b_t = smoothing trend in the value of t
a_t = smoothing alpha on the value of t
I_{SP} = seasonal index planting

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

F_t = prediction on the value of t
m = period
h = year forecasting
t = time

Note that the timing showing the difference with classic smoothing method which is used at the time of planting methods planting seasonal exponential smoothing index.

Algorithm 2 is the prediction for the period ahead as h for each cluster with exponential smoothing method of planting seasonal index. Seen on line 9, the prediction is done by using the value of smoothing alpha and smoothing beta multiplied by the value of the forecast period so that the trend component in the process of prediction is only an effect on any increase in the amount of the forecast period and the predictions will continue to do as long as the value of *i* as a reference the initial period ≤ value of *h* (line 7).

Algorithms 2 Prediction by planting seasonal exponential smoothing index

Input: $result[N] \leftarrow result\ fitting\ group\ rainfall\ as\ many\ as\ N$
1. $L \leftarrow the\ number\ of\ growing\ months\ (4\ months)$
2. $h \leftarrow the\ number\ of\ conducted\ production\ periods$
3. $alpha[N] \leftarrow alpha\ smoothing\ in\ every\ group$
4. $beta[N] \leftarrow beta\ smoothing\ in\ every\ group$
5. $index \leftarrow N$
6. $i \leftarrow 1$
7. **While** $i \leq h$ **then**
8. **For** $j = 1$ **to** L **do**
   $index \leftarrow index + 1$
9. $result[index] \leftarrow alpha[N-L+j] + cons.beta[N-L+j] * i$
10. **If** $j = L$ **then**
11. $i \leftarrow i + 1$
12. **End If**
13. **End For**
14. **End While**

C. Tests on Prediction Model of Exponential Smoothing Seasonal Planting Index.

The third phase of testing the model, testing was conducted to find the best prediction method to be used in the process of determining the planting season. Tests using the test size ME, RMSE, MPE, MAPE, MASE, and Euclidean Error to search results prediction error value, the smaller the error value the more accurate the prediction results.

Model predictions have differences with the actual value of the predictive value is generally called as a residual. The size of the samples used in determining the index of the prediction error is obtained from the value of the distance between the point of actual and predicted values point can be seen in equation 16 and equation 17. Assuming that the margin or error is obtained by using euclidean distance equation [23] :

$$d(p,q) = d(q,p) = \sum_{i=1}^{n} |(p_i - q_i)|^2 \qquad (16)$$

Where :
p = actual value / value of observation
q = prediction value / theoretical value

The percentage of the overall forecasting error is obtained by using the equation :

$$\frac{\sum d(p,q)}{\sum p^2} * 100 \qquad (17)$$

Algorithm 3 carry out the process to determine the index error reduction process (line 4) by looping in order to determine the difference or distance between the two points (actual point to point prediction or fitting). After that there is a process in determining the percentage of errors divider forecasting process based on the aggregate obtained (line 8).

Algorithm 3 Measurement Test of *euclidean error*

Input: $data[N] \leftarrow Actual\ data\ rainfall\ of\ N,$ $result[N] \leftarrow predicted\ result\ of\ N$
1. $error[N] \leftarrow the\ error\ value\ or\ difference$
2. $sigma[N] \leftarrow squared\ value\ of\ rainfallactual\ data$
3. $EE \leftarrow the\ percentage\ of\ forecast\ error$
4. **For** $i = 1$ **to** N **do**
5. $sigma[i] \leftarrow data[i]^2$
6. $error[i] \leftarrow |(data[i] - result[i])^2|$
7. **End For**
8. $EE \leftarrow \frac{sum(error)}{sum(sigma)} * 100$

Accuracy or precision of the method is a measure for determining the appropriateness or accuracy of the method is able to reproduce the data in the coming period, the smaller the error value is getting better and precise methods used.

Data from rainfall prediction algorithm Planting Exponential Smoothing Seasonal Index will be used to determine the planting season. The model can be used to determine cropping patterns because it is able to predict rainfall for the long term (minimum 12 months) with minimum error value.

IV. EXPERIMENTAL STUDIES

The experimental results and the evaluation of rainfall prediction models will be analyzed in section IV.

A. Initial Data Processing Experiment Research

Test field data was also conducted with respect to graphical visualization districts of Ngemplak rainfall data from 2003 to 2014 are presented in Figure 3. Noise is clearly visible on the chart with the missing value; one of the noises in the data in 2004, 2005, and 2006, in 2008, 2010 and 2014 is shown with a disruption in the line connecting the rainfall data.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

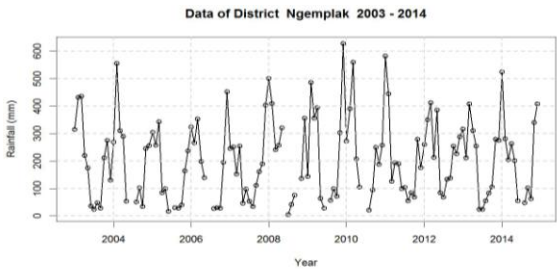


Fig. 3. Chart of rainfall data from 2003 - 2014

Rainfall data Chart of District Ngemplak from 2003 until 2014 after preprocessing is presented in Figure 4. Noise is not visible on the chart due to missing value has been filled using linear interpolation method, the data of rainfall in June and July of 2004 filled values 80 and 52 are shown from connecting line connect to the rainfall data.

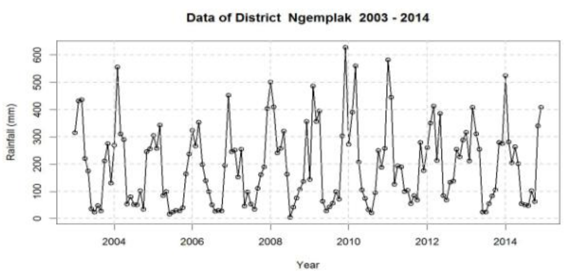


Fig. 4. The Chart of rainfall from 2003 to 2014 after preprocessing

The process of grouping data based SPI experiments conducted using rainfall data district Ngemplak in the period 2003 to 2014. The process begins by classifying the rainfall data into vectors u, v vector and the vector w. Each vector consists of four months; the vector u consists of data from January, February, March, and April for the period 2003 to 2014. The vector v consists of data in May, June, July, and August for the period 2003 to 2014. The vector w consists of the data in September, October, November, and December for the period 2003 to 2014. in the period 2003 to 2014 there were four groups, each group consisted of four months as shown in table 1.

TABLE 1
DATA RAINFALL IN DISTRICT NGEMPLAK BASED GROUP SPI

Year	Month											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
	u				*v*				*w*			
2003	315	432	437	220	174	35	24	47	28	211	275	130
2004	269	556	310	290	54	80	52	50	102	34	246	256
2005	305	258	344	85	100	17	24	30	28	40	165	237
2006	324	267	354	198	139	99	51	27	30	28	195	453
2007	245	252	152	255	46	98	54	34	111	162	190	404
2008	502	410	241	258	321	163	5	41	76	106	136	356
2009	143	486	357	395	64	28	43	57	100	71	304	628
2010	272	390	561	207	105	74	33	21	95	250	188	257
2011	583	445	126	194	189	99	104	55	85	68	280	176
2012	261	351	413	213	386	84	68	133	137	254	227	288
2013	316	212	409	310	255	24	24	55	83	105	280	275
2014	525	281	205	264	201	55	51	47	102	62	341	408

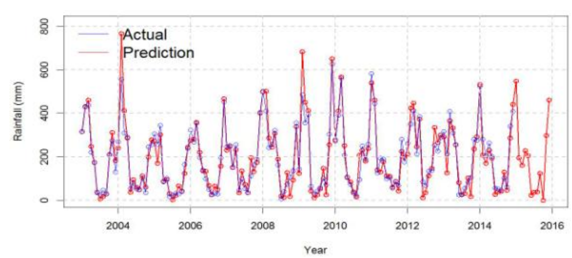
B. Experiments Rainfall Prediction Model of Exponential Smoothing Seasonal Planting Index

Exponential smoothing method is characterized by the determination of the parameters as the main parameter in the smoothing process. In this research smoothing parameter specified by planting seasonal index approach, shown in equation 18, with elections $I_{SP} = 3/4$,and the result.

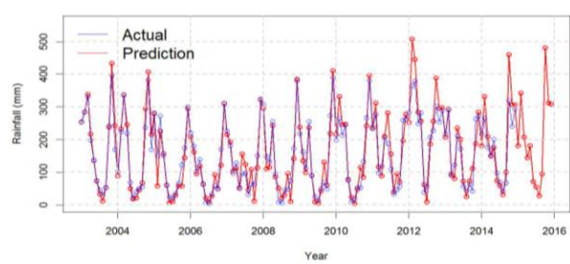
$$\alpha_{I_{SP}} = 1 - \exp(-I_{SP}) = 0.527633 \tag{18}$$

The value of smoothing parameter (I_{SP}) is the value of general or may not be equal to 3/4 if there are differences in the pattern of annual rainfall conditions and the difference one life cycle of the plant (the time from planting to harvesting.

The chart of rainfall prediction in 2015 for districts Ngemplak (Figure 5 (a)), sub Juwangi (Figure 5 (b)), and keacamatan Andong (Figure 5 (c)) shows the pattern of variation cycle prediction data from 2003 - 2014 follow previous period and in accordance with the pattern of variation of actual data cycle. Graphs show their predictions point is almost equal to the actual point in almost all lines of observation. Critical analysis of prediction experiments with planting seasonal exponential smoothing algorithm index is a prediction line 2015 shows a tendency seasonal pattern, the pattern indicates conformity with the pattern of historical data (actual data of previous years). Analysis of the data pattern predictions show that the algorithm seasonal planting appropriate index is used to predict the rainfall data for the planning of cropping pattern for seasonal patterns predicted results. The analysis also showed a trend of seasonal planting index algorithm suitable for the prediction of medium and long term.

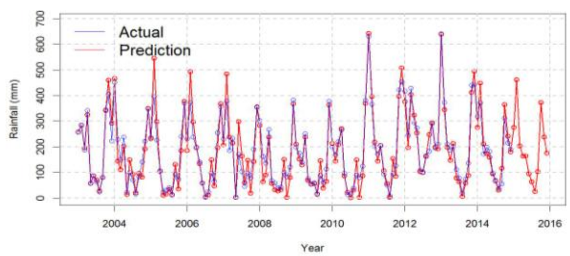


(a)



(B)

https://sites.google.com/site/ijcsis/
ISSN 1947-5500



(C)

Fig. 5. Predicted precipitation 2015 district. Ngemplak (a), district. Juwangi (b), and the district. Andong (c) using the method of planting seasonal index

C. Error Testing and Measuring Accuracy Prediction Model

The experimental results are planting seasonal prediction models index (Predict in the sample) rainfall by using the training data smoothing SPI value α = 0.527633 for all three districts showed good agreement. This is demonstrated the suitability of the data patterns of rainfall prediction in the original data. Table 2 shows the value ME, RMSE, MAE, MPE, MAPE, and EE relatively small. Testing accuracy by smoothing SPI value α = 0.527633 for districts Ngemplak generate ME value of 1.30, RMSE of 44.93, MAE at 30.78, at 27.44 MPE, MAPE at 27.44, and EE (Euclidean Error) amounting to 3.38. Predict out of the plot of the sample can be seen the difference value data and the predictive value of future rainfall data in accordance with the pattern of the original data.

TABLE 2
TESTING ERROR PREDICTION USING PROPOSED ALGORITHMS

No.	District	Test Measurement	Proposed Algorithms
1	Ngemplak	ME	1.30
		RMSE	44.93
		MAE	30.78
		MPE	3.74
		MAPE	27.44
		EE	3.38
2	Juwangi	ME	4.79
		RMSE	37.32
		MAE	26.93
		MPE	12.46
		MAPE	35.19
		EE	3.96
3	Andong	ME	1.07
		RMSE	40.56
		MAE	29.33
		MPE	3.96
		MAPE	23.73
		EE	3.24

The level of accuracy of rainfall prediction excl. Ngemplak 2015 using exponential smoothing algorithm planting seasonal smoothing SPI index with a value of α = 0.527633 based on the size of the test euclidean error of 96.62%. Data rainfall prediction results 2015 to kec. Ngemplak, excl. Juwangi, and excl. Andong showed a tendency seasonal pattern, the pattern indicates conformity with the pattern of historical data. Experiments proved that the result of rainfall prediction by planting seasonal exponential smoothing algorithm index can predict the long-term rainfall data with a good degree of accuracy and data prediction results have seasonal data pattern so that proper planning is used for rice cropping pattern.

Comparison of measurement error between classic Exponential Smoothing (ES) algorithm with seasonal planting for fitting index of rainfall data from 2003 to 2014 and forecast rainfall data of 2015 are presented in Table 3. The analysis shows seasonal planting index algorithm predominantly (5 of 6 size test) has an average value of error is much smaller than the classic exponential smoothing algorithm. The accuracy of prediction algorithms planting seasonal index by 95.73% better than the exponential smoothing algorithm α = 0.1 by 56.55%, and exponential smoothing of α = 55.53.

TABLE 3
COMPARISON OF THE AVERAGE ERROR VALUE BETWEEN THE EXPONENTIAL SMOOTHING ALGORITHM AND PROPOSED

No.	Test Measurement	E.S. AlgorithmsWith ∝ = 0.1	E.S. AlgorithmsWith ∝ = 0.2	Proposed Algorithms
1	ME	1.17	1.49	5.17
2	RMSE	163.74	166.10	51.37
3	MAE	132.09	129.26	35.19
4	MPE	291.42	243.95	32.05
5	MAPE	326.98	284.85	56.25
6	EE	43.45	44.75	4.27

The results of the test measurement prediction of rainfall data in Boyolali Regency using algorithms seasonal planting index for ME value of 5.17, RMSE values of 51.37, MAE value of 35.19, MPE value of 32.05, the value of MAPE by 56 , 25, and EE value of 4.27.

Comparison of test measurement values between classic exponential smoothing algorithm with the proposed model for the 19 districts in Boyolali Regency in Table 3, show that the proposed models have value better prediction accuracy. The proposed model is better because it has a value of the average error is smaller than the prediction model with exponential smoothing. The average increase in the accuracy of rainfall prediction in Boyolali Regenecy, Central Java Province, Indonesia using seasonal planting index algorithm is 40.2%.

From the analysis of the results of raifall prediction experimental modeling using algorithm Exponential Smoothing classic and algorithms Exponential Smoothing Using Seasonal Planting Index (ESSPI) it can be concluded that the value of the test error for the exponential smoothing classic is big with a level of accuracy is low and the data predicted results rainfall data for long term show trend pattern. While the value of the test error for exponential smoothing seasonal planting index algorithm is small (close to 0), with a high accuracy rate of 95.73%, and the data predicted results for long-term rainfall data showed a seasonal pattern. In this research, data of rainfall prediction will be used to determine cropping patterns that required

the model to predict rainfall for long term (minimum 12 months) with minimum error value. Recommendations based on the analysis of experimental results, rainfall prediction model that is appropriate for the needs of planning cropping is Exponential Smoothing Seasonal Planting Index Algorithm (ESSPI).

Review journal literature and research on predictive models that have been implemented in this research indicates novelty because it has not been studied or discussed by other researchers. Novelty of this research is the methods and new algorithms for clustering rainfall data based on seasonal planting index, new methods and new algorithms to determine the value of smoothing (α) algorithm based on seasonal planting index, and new algorithms rainfall prediction based exponential smoothing seasonal planting index. Overall algorithms and models will be implemented in the form of a new package of R programming language that can be used widely.

V. CONCLUSION

Rainfall prediction model uses exponential smoothing seasonal planting index algorithm shown to predict short-term and long-term predictions with good accuracy consistently, the average accuracy rate of 95.73%. Rainfall prediction by a combination of exponential smoothing seasonal planting index has a value of prediction accuracy is better and has a value of error is relatively smaller than the classic exponential smoothing, the average increase in the accuracy of rainfall prediction Boyolali Regency, Central Java Province, Indonesia is 40.2% , From the measurement accuracy and measurement error can be concluded that the prediction of rainfall data for 19 sub-districts in the Boyolali Regency for the prediction of 12 periods exponential smoothing algorithms using seasonal planting index eligible to be used in the preparation of the planting season.

Future studies are recommended to expand the scope of the study area (Central Java area or all provinces in Indonesia). The addition of variables determining the cropping pattern with the soil condition data and the data of plant varieties. Conduct a comparative study on the research methods of prediction with the prediction method of neural networks and spatial regression kriging interpolation method.

ACKNOWLEDGMENT

The author wishes to thank the Department of Agriculture Forestry and Plantation of Boyolali Regency, Central Java Province Indonesia that Provided the data rainfall for the analysis.

REFERENCES

[1] Boer, R., A. Buono, Sumaryanto, E. Surmaini, A. Rakhman, W. Estiningtyas, K. Kartikasari, and Fitriyani, “Technical Report on Vulnerability and Adaptation Assessment to Climate Change for Indonesia’s Second National Communication”, Ministry of Environment and United Nations Development Programme, Jakarta, 2009.

[2] Bappenas, 2010, diakses dari http://www.bappenas.go.id/files/2613/ 5183/0762/sektor-pertanian__20110217190317__0.pdf pada 15 Januari 2014.

[3] Ngopya, F., “The Use Time Series in Crop Forecasting”, Regional Early Warning System for Food Security, Food, Agriculture and Natural Resources (FANR) Directorate, Botswana, 2009.

[4] Burkom, S.H., Muphy, S.P., dan Shmuelli, G., “Automated Time Series Forecasting for Rainfall”, The Johns Hopkins University Applied Physics Laboratory, Department of Decision & Information Technologies, Robert H. Smith School, University of Maryland CollegePark, 2006.

[5] Makridakis, S., Victor, E., & Steven, C., “Forecasting Methods and Applications”, Binarupa Aksara Publisher, Jawa Barat, 2002.

[6] Naylor, R.L., Battisti, D.S., Vimont, D.J., Falcon, W.P., dan Burke, M.B., “Assessing risks of climate variability and climate change for Indonesian rice agriculture”, Proceeding of the National Academic of Science 114, pp. 7752-7757, 2000.

[7] Surmaini, E., Runtunuwu, E., dan Irsal, L., "Upaya Sektor Pertanian Dalam Menghadapi Perubahan Iklim”, Balai Besar Penelitian dan Pengembangan Sumberdaya Lahan Pertanian, Jalan Ir. H. Juanda No. 98, Bogor 16123, 2010.

[8] Mara, M.N., Satyahadewi, N., dan Yundari, “Kajian Teoritis Hybridizing Exponential Smoothing dan Neural Network Untuk Peramalan Data Runtun Waktu”, Bimaster Ilmiah Mat. Stat. Dan Terapannya (Bimaster), Vol. 2 No. 3, pp. 205-210, Pontianak, 2013.

[9] Ghil M., Allen M.R., Dettinger M.D., Ide K., Kondrashov D., Mann M.E., Robertson A.W., Saunders A., Tian Y., Varadi F., &Yiou P., “Advanced Spectral Methods For Climatic Time Series”, Reviews of Geophysics, 40, pp. 1-1 – 1-41, 2002.

[10] Ai, T.J., “Optimalisasi Prediksi Pemulusan Eksponensial Satu Variabel Dengan Menggunakan Algoritma Non Linear Programming”, Jurnal Teknologi Industri 3 (3), pp. 139 – 148, 1999.

[11] Raharja, A., Wiwik, A., &Retno, A.V., “Penerapan Metode Exponential Smoothing untuk Peramalan Penggunaan Waktu Telepon Di PT. Telkom Divre 2 Surabaya”, Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh November, SsISFO-Jurnal Sistem Informasi, 2009.

[12] Unkel, S. & Farrington, C.P., “Statistical methods for the prospective detection of infectious disease outbreaks: A review”, Department of Mathematics and Statistics, Faculty of Mathematics, Computing & Technology, The Open University, Walton Hall, Milton Keynes, UK, 2010.

[13] Lu, H.M., Zeng, D., & Chen, H., “Prospective Infectious Disease Outbreak Detection Using Markov Switching Models”, Management Information Systems Department, University of Arizona, Tucson, 2008.

[14] Dong, Z., Yang, D., Walsh, M.W., Reindl, T., dan Aberle, A., “Short-term Solar Irradiance Forecasting Using Exponential Smoothing State Space Model”, Solar Energy Research Institute Of Singapore, National University of Singapore, Singapore, January 2013.

[15] Sofla, M.J., Silahi, B., dan Masomi, M. T., “Germi County Seasonal Precipitation Routing and Analysis, Using Holt-Winter Method for Times with Non-seasonal changes”, Technical Journal of Engineering and Applied Sciences, Vol 3(11): 950-953, 2013.

[16] Mahsin, M., “Modelling rainfall in Dhaka Division of Bangladesh Using Time Series Analysis”, Journal of Mathematical Modelling and Application, Vol 1. No. 5, pp. 67-73, 2012.

[17] Gunawan, D. Dan Linarka, U.A., “Penentuan Prediktor Untuk Prediksi Curah Hujan Bulanan Menggunakan Metode Statistical Dynamical Downscaling”, Jurnal Meteorologi dan Geofisika, Vol 12. No. 1, pp. 93-102, 2011.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

[18] Swarinoto, W.S., "Evaluasi Kehandalan Simulasi Informasi Prakiraan Iklim Musiman Menggunakan Metode ROC", Jurnal Meteorologi Dan Geofisika Volume 11 Nomor 2 : 116 – 126, 2010.

[19] Swarinoto, W.S. dan Makmur, E.E.S., "Simulasi Prediksi Probabilitas Awal Musim Hujan dan Panjang Musim Hujan", Buletin Meteorologi Klimatologi dan Geofisika, Vol. 5 No. 3, pp. 340 – 353, 2012.

[20] Tresnawati, R., Nuraini, A.T., dan Hanggoro, W., "Prediksi Curah Hujan Bulanan Menggunakan Metode Kalman Filter Dengan Prediktor SST Nino 3.4 Diprediksi", Jurnal Meteorologi Dan Geofisika Volume 11 Nomor 2 : 106 – 115, 2010.

[21] Setyono, A., dan Novianto, S., "Penerapan Interpolasi Linier Untuk Deteksi Garis Lurus Pada Citra", Techno.COM, Vol. 12 No. 3, pp. 143 – 149, Agustus 2013.

[22] Prasetyo, S.Y.J.P., Subanar, Winarko, E. & Daryono, B.S., "The Prediction of Population Dynamics Based on the Spatial Distribution Pattern of Brown Planthopper (Nilaparvata lugen Stal.) Using Exponential Smoothing – Local Spatial Statistics", Journal of Agricultural Science, published by Canadian Center of Science and Education, 5(5):1916-9752, 2013.

[23] Nugraheny, D., "Metode Nilai Jarak Guna Kesamaan atau Kemiripan Ciri Suatu Citra (Kasus Deteksi Awan Cumulonimbus Menggunakan Principal Component Analysis", Jurnal Angkasa, Vol. VII No. 2, pp. 21 – 30, November 2015.

[24] Naik, B.R., Latha, T.H., dan Babu, C.H., "Command Area Development by Using FAO Cropwt 8.0 Model and Impact of Climate Change on Crop Water Requirement", International Journal Of Applied Research, Vol. 1 No. 13, pp. 142 – 155, November 2015.

[25] Ishaque, M., and Ziblim, S.D., "Use of some exponential smoothing models in forecasting some food crop prices in the upper east region of Ghana", Journal of Mathematical Theory and Modeling, 3: 16-27, 2013.

[26] Jalil, N.A., Ahmad, M.H., and Mohamed, N., "Electricity load demand forecasting using exponential smoothing methods", World Applied Sciences Journal, 22 (11): 1540-1543, 2013.

[27] Meesad, P., and Srikhacha, T., "Adaptive approach and seasonal technique for universal data forecasting", Reseach paper : Department of teacher training in electrical engineering, faculty of technical education, KMITNB, 2006.

[28] Susanti E. F. R , June T., and Amien L.I., "Utilization of Climate Information for Development of Early Warning System for Brownplanthoppers Attack on Rice", Indonesian Hydrology and Agroclimate Research Institute, JalanTentaraPelajar No. 2, Bogor 16111, West Java,Indonesian Journal of Agriculture 3(1), 2010:17-25, 2010.

[29] Dominiak A.W. &Zeug-Żebro K., "Spatial statistics in the analysis of county budget incomes in Poland with the R CRAN", Proceedings of 30th International Conference Mathematical Methods in Economics, 2006.

[30] Ma, X. & Pei, T., "Exploratory Spatial Data Analysis of Regional Economic Disparities in Beijing During 2001 – 2007", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 38(2):241-244, 2009.

AUTHORS' INFORMATION

Kristoko Dwi Hartomo, Received is B.Sc. at Informatics Engineering from the Duta Wacana Christian University Yogyakarta Indonesia. Received is Master degree at Computer Science at Faculty of Mathematics and Natural Sciences from the Gadjah Mada University Yogyakarta Indonesia.He is student Doctoral Program Computer Science at Faculty of Mathematics and Natural Sciences from the Gadjah Mada University Yogyakarta Indonesia. He is currently a lecturer in the Informatic Engineering Department, Faculty of Information Technology, Satya Wacana Christian University Salatiga Indonesia.His current research interests include spatial Modeling, GIS, database, data mining and their applications.

Subanar, Received the M.Sc. degree in Statistics from Faculty of Mathematics and Natural Sciences GadjahMada University Yogyakarta Indonesia in 1971. He received his Ph.D. degree in Statistics from the University of Wisconsin United States in 1976. He is currently a lecturer and professor in the Faculty of Mathematics and Natural Sciences from the GadjahMada University Yogyakarta Indonesia. His current research interests include Mathematic Statistics, Neural Network and Bootstrap Method.

Edi Winarko, Received the B.Sc at Statistics from Faculty of Mathematics and Natural Sciences GadjahMada University Yogyakarta Indonesia in 2001, M.Sc. degree in Computer Science from the Queen's University Canada, United States in 2002. He received his Ph.D. degree in Computer Science from the Flinders University of South Australia in 2007 He is currently a lecturer in the Faculty of Mathematics and Natural Science from the GadjahMada University Yogyakarta Indonesia. His current research interests include data warehousing, data mining, and information retrieval.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 6, June 2016

A New MultiPathTCP Flooding Attacks Mitigation Technique

Adwan Yasin
Department of Computer Science
Arab American University
Jenin, Palestine

Hamzah Hijawi
Department of Computer Science
Arab American University
Jenin, Palestine
.

***Abstract*—MPTCP is a new protocol proposed by IETF working group as an extension for standard TCP, it adds the capability to split the TCP connection across multiple paths. It provides higher availability and improves the throughput between two multi-address endpoints. Many Linux distributions have been developed to support MPTCP, most of them are open source which can be modified and compiled to support different experimental scenarios. Splitting the single path TCP connection across multiple paths adds new challenges in paths management and raises new security threats. Some of these threats include flooding and hijacking attacks performed by on-path and off-path attackers. In this article, we propose a new algorithm to mitigate the flooding and hijacking attacks in MPTCP, the proposed method allows a stateful processing of the initial SYN message and it's following SYN_JOIN messages.**

Keywords*—*TCP, MPTCP, flooding, hijack, on-path, off-path, flooding, DoS

I. INTRODUCTION

TCP is the most transport protocol used on the internet today, it has been used by most applications as a reliable transfer protocol to transfer data between endpoints. TCP first design was in the 1970s, it has been evolved and enhanced to the current design we have today. TCP was implemented as a layer four protocol in the OSI model stack and as a design decision, the separation from the network layer is intended to be hidden. Five tuples are used to distinguish different streams from each other and to demultiplex packets to their appropriate destination. Source and destination IP addresses are used to forward the pack from source point to the destination, source and destination port numbers are used to identify the running processes on the source and destination while protocol identifier is used to indicate that the connection is using TCP. Therefore, any TCP connection is bounded to a unique socket through a single path between two endpoints [1]. However, if one of the five tuples is changed after the connection is established then the connection will fail.

The design of networks is changed, servers are becoming multi-homed, data centers have many redundant links and mobile devices have multiple wireless interfaces [2]. In order to make a use of these redundant connections, a new TCP design was evolved, it is called multipath TCP [3]. MPTCP allows multiple paths if they exist between the two communicating hosts to be effectively and concurrently used by a single TCP connection. MPTCP has obvious benefits for availability, reliability and load balancing [3]. It is more robust and can achieve better performance compared with a single-path TCP. One of the primary MPTCP design goals is maintaining the compatibility with existing applications and network infrastructure. This is achieved by presenting the MPTCP as a sublayer under TCP layer and let the TCP handles the upper layers applications [4, 5].

MPTCP connection consists of one or more TCP connection. Thus, the risk of vulnerabilities exist in MPTCP would be at least of the same risk in TCP, and particularly the attacks which performed by an on-path attacker who may impersonate one of the communicating parties and eavesdropping, forging, dropping or hijacking the session [6]. One of the design goals of MPTCP is that it is should at least perform as the standard TCP. So, the set of new vulnerabilities exist from the capability of adding new paths to an ongoing connection must be explored. Mainly, flooding and hijacking attacks which are performed by off-path and on-path attackers, and can result in redirection the traffic to unintended target [6, 7].

This paper addresses the flooding and hijacking attacks on MPTCP and proposes a new solution to mitigate these types of attacks. The rest of the paper is organized as follow. Related work is provided in section 2, section 3 gives an overview about MPTCP. Connection establishment in MPTCP is explained in section 4. In section 5, multiple flooding attacks scenarios are explained. The hijacking attack is described in section 6. The proposed solution to mitigate these attacks is provided in section 7 and conclusions are discussed in section 8.

II. RELATED WORK

MPTCP is a new approach towards efficient load balancing between endpoints participating in the TCP connection, it was implemented in many Linux-based distributions [8]. As a design decision, MPTCP is totally backward compatible with existing applications and network devices. A comprehensive study on the impacts that the protocol may have on TCP applications was summarized in [9] and the compatibility issues between MPTCP and standard TCP have been discussed. A performance analysis of MPTCP

have been made in [10], in which, throughput comparisons were made between standard TCP and MPTCP with different scenarios, the experiments show how MPTCP outperforms standard TCP in terms of throughput and handover capability when the connection lost. MPTCP offers benefits for availability and connectivity, but there is also a security risk which must be addressed. One of the potential risks comes from the traffic fragmentation between the different paths between the two endpoints. However, modern network security technologies like IPS and IDS are not ready for MPTCP, they are not currently able to re-assemble a full MPTCP session from the different paths and properly inspect and represent a potential security risk [11].

A threat analysis for MPTCP is provided in RFC6181 [12], the analysis identified and characterized the new vulnerabilities which may appear after supporting multiple paths in a single TCP connection. As one of design goals of MPTCP, it is assumed that any MPTCP connection should at least perform as a single path TCP, this means that any potential risk in TCP must be addressed in MPTCP. The studies in [6, 12] provided analyses about the most common potential threats which may exploit the MPTCP connection, this includes flooding and hijacking attacks. A basic solution to mitigate these attacks were provided in [6], in which the sender asks the receiver for each new sub-flow if it can accept data from this new connection If yes then they exchange a random token for authentication purpose. The architecture design for MPTCP provided in [13] suggested three key security requirements, MPTCP should be able to provide a mechanism to confirm that the endpoints participating in a sub-flow handshaking are the same endpoints in the original connection establishment. MPTCP should also provide a mechanism to verify that a host can receive traffic at a given address before opening the sub-flow, it should also provide replay protection in order to verify that a request to add or remove a sub-flow is fresh.

III. MPTCP OVERVIEW

Today's networks are becoming multipath, most of the servers, data centers, and mobile devices have redundant network interfaces and more than one IP address at the same time. MPTCP was designed to utilize all available paths between the two communicating points [3]. Figure 1 shows MPTCP connection for a mobile device which has two network interfaces. WiFi is the main connection and 3G is the backup one.

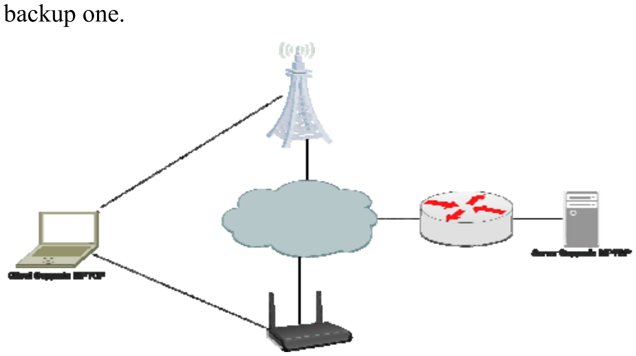


Figure 1. MPTCP topology

MPTCP was designed to achieve a set of requirements which are summarized in three main design goals. The first one is improving the throughput compared with a single-path TCP. The second one, do not harm; MPTC should not take capacity more than a standard TCP would take if both share the same path. The last goal is balancing the congestion; MPTCP should move the traffic to the least congested paths [4, 6]. Two design decision were taken in consideration in MPTCP implementation; application and network compatibility. Application compatibility means that MPTCP should work with existing applications running with TCP without any modification and the network compatibility means that MPTCP should operate with existing networks [3]. As a result of these two design decisions, MPTCP is implemented as a sub-layer in the transport layer, and this implementation achieves the transparency of existing multiple paths to the upper layers as shown in figure 2.

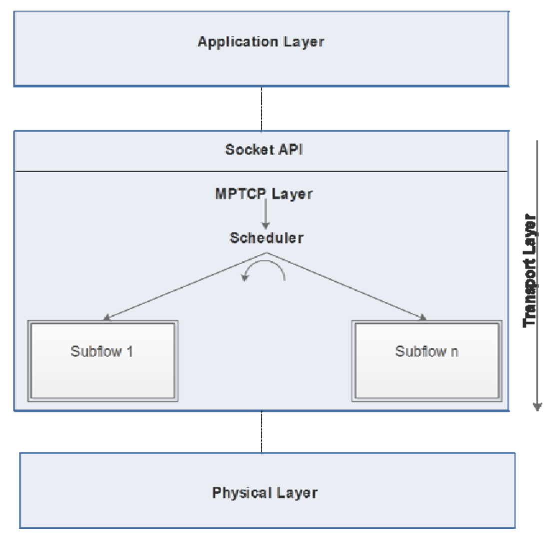


Figure 2. MPTCP in TCP/IP stack

For each path between the source and destination, a new sub-flow is created, each sub-flow can be considered as a normal TCP connection and can be distinguished with the five tuples. Scheduler is a part of MPTCP implementation, it is used to schedule the traffic between all related sub-flows [15]. With the possibility of using multiple paths between the source and destination, concerns have arisen about congestion control over these paths. However, congestion control in MPTCP is different from standard TCP. One of the requirements for MPTCP congestion control algorithm is to be fair to the standard TCP if both share the same link. Another requirement is to transfer more traffic to the least congested path [10], this requirement is needed to utilize the paths between the source and the destination as much as possible. However, if one of the paths is congested then MPTCP decreases the window size on this path and increases the window on the least congested paths [16].

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

IV. MPTCP CONNECTION ESTABLISHMENTS

Standard TCP connection can be divided into three stages; connection establishment, data transfer, and connection release. Connection establishment starts with a three-way handshake. However, in order to open a connection, the client sends a synchronize request SYN to a port in which the server is listing. All connection relevant information are sent in SYN request, this includes the source port and the initial sequence number. The server then acknowledges the SYN request with SYNACK reply message. After that, the client acknowledges the SYNACK and then the connection is established. The connection is now established and both hosts can start sending data packets. After the data transfer is over, the connection must be closed, this happens by using FIN packets, the connection is terminated after the FIN packet is acknowledged by both hosts [3].

Multipath TCP connection is established in the same way as TCP connection is established, it uses a three-way handshakes and the options field in the TCP header. The mp_capable option is set in the SYN packet to indicate that the source can perform MPTCP. The destination then replies with SYNACK packet, if it also supports MPTCP then MP_CAPABLE is set, the source then replies with ACK packet which has the MP_CAPABLE option to ensure that this is a MPTCP connection. After the connection is established, participating hosts can add new sub-flows to the connection by using the same negotiation procedure applied in the connection establishment. MP_JOIN option is used instead of MP_CAPAPLE with the connection identifier to inform the destination which connection it would like to join [17]. Once the connection has multiple sub-flows, scheduler decides how to distribute the traffic between them. Each sub-flow can be considered as a standalone TCP connection which has its own congestion control algorithm and sequence numbers space. A new sub-flow connection can be established and added if a new path is available and can be removed if the path is vanished [18]. MPTCP connection establishment is summarized in figure 3.

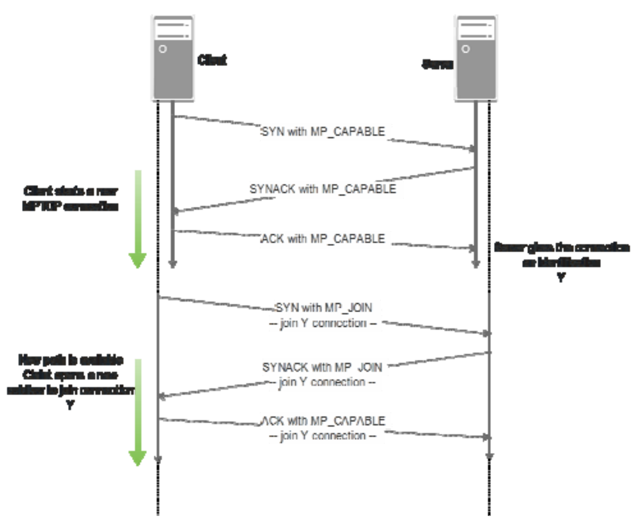


Figure 3. MPTCP connection establishment

V. MPTCP FLOODING ATTACKS

MPTCP flooding attack is one of the attacks introduced by address agility, the goal from this attack is to exhaust the victim by a heavy traffic causing a denial of service. Figure 4 illustrates the redirection attack in which the attacker uses a streaming server to redirect a huge amount of traffic to the victim host.

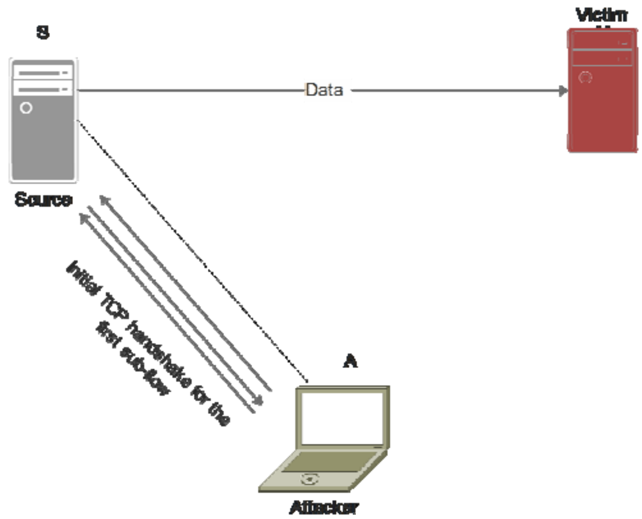


Figure 4. Flooding attack using a stream server

First, the attacker opens a MPTCP connection with the traffic source S and starts downloading a heavy traffic, this connection involves the IP addresses of the attacker and the server S. While the heavy traffic is coming to the attacker from S, the attacker adds the victim IP as one of the available addresses for the connection. After this step, the connection has two IP addresses for the attacker A and a single IP address for the traffic source S. The attacker goal at this point is to send the heavy traffic load from source S to the victim node V. To achieve that, the attacker pretends that the path between him and the source is congested while the path between the traffic source S and the victim V is not. As a result, most of the traffic will be shifted to the least congested path between S and V. In order to successfully complete this step, the attacker acknowledges the traffic that flows between S and V and does not acknowledge the traffic that flows between S and A. ACKs must be sent using packets contain the IP address of the victim as a source address. Sequence numbers of the data being transmitted between S and V should also be known by the attacker. Once the attacker manages to send ACKs in path between S and V, the traffic will start hitting the victim machine while source S thinks it is sending the traffic to A. In order to increase the amount of the traffic hitting the victim, the attacker needs to increase the windows size for the path between S and V, in addition to simulate the congestion in the path between the source and the attacker nodes.

The effect of this type of flooding attacks can be significantly increased if the attacker uses more than one streaming server at the same time causing a distributed attack. However, the attacker can repeat the previous scenario with

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

many servers causing the traffic to be redirected from multiple servers at the same time, as shown in figure 5.

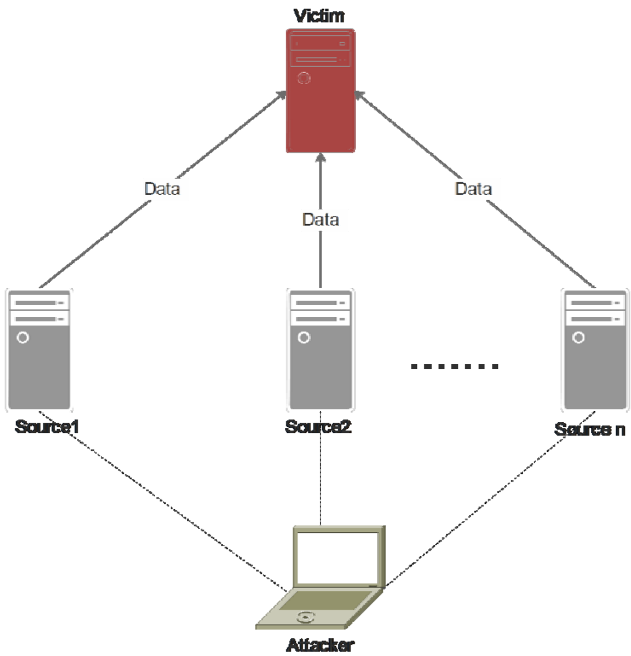


Figure 5. Flooding attack using multiple stream servers

Another type of flooding attacks is MPTCP SYN flooding attack. This attack uses the SYN messages in order to exhaust the victim resources and prevents new sub-flows connections [19]. The attacker starts with a normal MPTCP session by sending regular SYN packet and then sends many MP_JOIN requests as supported by the server, each join message is sent with different source IP and source port combinations. This is an amplification attack, in which the cost on the server side is the cost needed for the initial SYN request in addition to the cost needed for all following SYN MP_JOIN requests. Figure 6 illustrates this attack, the attacker uses a list of N IP addresses to open one MPTCP connection with N-1 sub-flows.

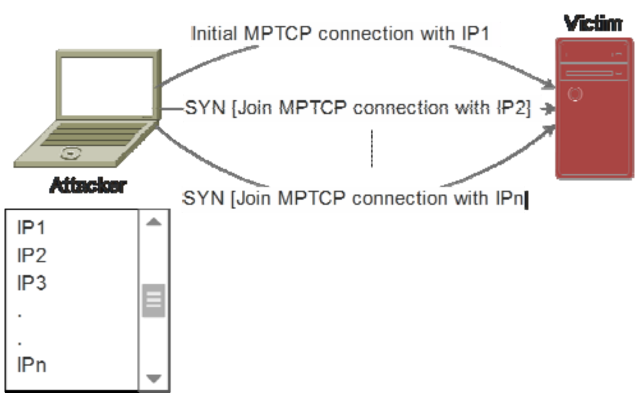


Figure 6. Single connection MPTCP SYN flooding attack

In order to increase the effect of SYN flooding attack, the attacker can use each IP in the list to open a new MPTCP connection instead of joining an existing one, the rest of IP addresses can be used with different ports combinations to join the connection. In this case, the attacker can open N MPTCP connections with N-1 sub-flows as shown in figure 7.

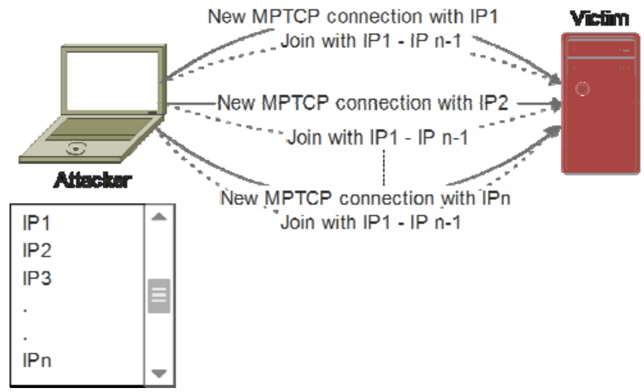


Figure 7. Multiple connections MPTCP SYN flooding attack

VI. MPTCP Hijacking Attacks

In this type of attacks, the attacker attempts to hijack the MPTCP connection in order to personate one of the legitimated peers. It happens after the initial MPTCP connection is established and the two peers are exchanging data. The target from this attack is either eavesdropping or altering the data being transferred between the two peers. Figure 8 shows the general overview of MPTCP hijacking attack.

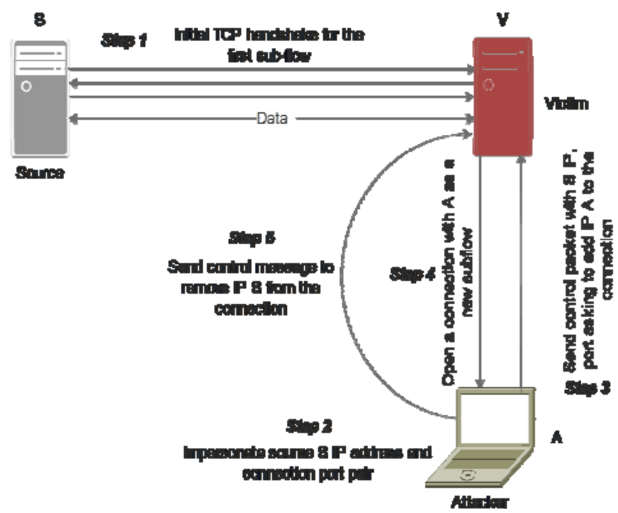


Figure 8. MPTCP hijacking attack

After the connection is established in step 1, the attacker needs to figure out the fourth tuples used to distinguish this connection. This information is needed in order to send a fake control packet which asks the victim machine to open a new sub-flow with the attacker. This request is sent by using S IP address and port number and it includes IP A in ADD_ADDR field in the MPTCP header. Since the request is sent by using source S information, the victim thinks it is a legitimated

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

request so it opens a new sub-flow with the attacker. Now the connection has two sub-flows, the first one between the source and the victim and the second one between the victim and the attacker. At this point, the traffic started to be split between these two connections. In order to complete the attack, the attacker sends another control message to remove IP S from the list of addresses related to the connection. After this step, the hijacking attack is successfully completed and the traffic starts flowing between the victim and the attacker. The attacker may modify or just eavesdropping on the data and then forward it to the source machine. In order to keep the connection alive, the attacker repeats the same procedure with the source machine which makes both peers think they are talking to each other while they are talking to the attacker.

VII. Proposed Algorithm

The idea behind the proposed algorithm is mainly based on three actions. First, we enforce the client by the implementation to send all relevant information about all paths that may come up while the connection is active, we called this information a Metadata. Information includes the physical interface MAC address and possible IP address for the path. This information if it is available will give the server an indication about future sub-flows that may come up. Second, limiting the maximum number of sub-flows for each MPTCP connection, this is already implemented in most Linux distributions [4, 20], we suppose that the maximum limit is five. The third one is using a hash key value to authenticate each sub-flow before allocating the resources on the server side. Suppose there are two hosts A and B, A wants to start MPTCP connection with B. Following steps must be followed in order to mitigate the flooding and hijacking attacks in the connection between A and B.

1. In order to start MPTCP connection, A sends SYN packet with MPTCP_Capable.
2. A includes in the SYN packet the Metadata information about all its candidate sub-flows (interfaces MAC address and possible IP addresses).
3. When B receives the first SYN packet, it stores temporarily the information related to this connection with all candidate sub-flows.
4. B replies with SYNACK packet with a crypto hash key generated related to the connection; the hash key is generated from the Metadata related to this sub- flow in addition to a random value chosen by the server.
5. B also includes a random key in the message. This key is necessary to eliminate the hijacking attacks; each future request to add a new sub-flow will have this key.
6. A stores the random key related to this connection.
7. B stores the hash key and the random key in a table related to this connection as shown in table1.

Table 1. Sub-flows hash table

Connection: Y, Random key: randKey	
Hash value	Is connection active
Hash1	Yes
Hash2	No
Hash3	No
Hash4	No
Hash5	No

8. When A receives SYNACK packet, it responses with ACK packet.
9. A should include the same hash key in the ACK packet.
10. When the ACK packet is received by B, it checks for the hash value.
11. If the hash value is the same as the one which was sent by SYNACK and the value of “is connection active” is no, then this connection is validated and “is connection active” value is set to yes.
12. Now, suppose there is a new sub-flow exists.
13. A sends SYN_JOIN packet to B, it includes the random key obtained in step 6 in the request.
14. B checks the table related to this connection and validate the random key. If it is validated, B continues with next steps.
15. B repeats the same authentication steps mentioned previously in steps 4, 8-11.
16. B checks if the new sub-flow information exists in the Metadata for this connection.
17. If yes, and the new sub-flow is authenticated, then B will add the new sub-flow to the connection.

For the case of the flooding attack described in figures 4 and 5, when the attacker starts a MPTCP connection with the streaming server, Metadata information is sent with SYN packet. If the attacker informs the server that it has a new IP address and it wants to start a new sub-flow, server checks if this IP exists in the Metadata. If not, then the server will immediately ignore the request. If it exists, then the server will send the crypto hash key to authenticate the new sub-flow. This process is repeated for each new sub-flow request. The algorithm grantee that only traffic requested by the host can reach it.

For the scenarios described in figure 6 and figure 7, the attacker wants to perform SYN flooding attack on the victim machine, the maximum number of allowed sub-flows for any MPTCP connection is assumed to be five. The attacker starts with SYN packet which contains MP_CAPABLE option, it forced by the implementation to send Metadata information about all possible sub-flows. When the victim receives the SYN packet, it generates a crypto hash key for the first five sub-flows and stores the values in a table as shown in table1. When the attacker attempts to perform a SYN flooding attack

by sending multiple SYN JOIN packets to the victim, for each request the victim calculates the crypto hash key, if the value exists in the table then this request could be legal. The victim continues with authentication process by sending SYNACK packet to the attacker with the hash related to the sub-flow merged with a random hash key. If attacker replies with ACK which contains the same hash key, then the sub-flow request is authenticated and added to the connection paths. The resources related to the sub-flow are only allocated after the sub-flow is validated.

For the case of hijacking attack described in figure 8, when the attacker sends a control packet to add its IP address to an existing MPTCP connection, it should include the random key obtained in the first MPTCP connection initialization as described in step 5 in the proposed algorithm. Since the attacker didn't start the connection it will not have the random key and will not be able to send a valid request to open a new sub-flow with its IP address.

VIII. CONCLUSION

Supporting multipath over TCP is the most significant change happens to TCP since the first design in the 1970s. It allows the traffic related to a single connection to be split over multiple paths which in term improves the reliability and increases throughput. Due to the address agility provided by MPTCP, new security threats appears, this includes flooding and hijacking attacks. In this article, we analyzed multiple flooding and hijacking attacks scenarios which may occur in any MPTCP connection, we also provided a proposed solution to mitigate these types of attacks.

References

[1] Postel, Jon. "RFC793: Transmission Control Protocol. USC." Information Sciences Institute 27 (1981): 123-150.

[2] Wischik, Damon, Costin Raiciu, Adam Greenhalgh, and Mark Handley. "Design, Implementation and Evaluation of Congestion Control for Multipath TCP." In NSDI, vol. 11, pp. 8-8. 2011.

[3] Bonaventure, Olivier, Mark Handley, and Costin Raiciu. "An overview of Multipath TCP.”; login: 37, no. 5 (2012): 17-23.

[4] Ford, C. Raiciu, and M. Handley. "TCP Extensions for Multipath Operation with Multiple Addresses" draft-ietf-mptcpmultiaddress (work in progress)." (2010).

[5] Scharf, Michael, and Alan Ford. Multipath TCP (MPTCP) application interface considerations. No. RFC 6897. 2013.

[6] Díez, Javier, Marcelo Bagnulo, Francisco Valera, and Iván Vidal. "Security for multipath TCP: a constructive approach." International Journal of Internet Protocol Technology 6, no. 3 (2011): 146-155.

[7] Eddy, Wesley M. "SYN Flood Attack." In Encyclopedia of Cryptography and Security, pp. 1273-1274. Springer US, 2011.

[8] "MPTCP Linux Kernel Implementation." MultiPath TCP. Accessed May 01, 2016. http://mptcp.info.ucl.ac.be/.

[9] Scharf, Michael, and Alan Ford. Multipath TCP (MPTCP) application interface considerations. RFC 6897, March, 2013.

[10] Hijawi, H. M., & Hamarsheh, M. M. Performance Analysis of Multi-path TCP Network. International Journal of Computer Networks & Communications (IJCNC), 8(2), 145-157.

[11] "The Dangers and Promise of Multipath TCP." The Dangers and Promise of Multipath TCP. Accessed May 01, 2016. http://www.enterprisenetworkingplanet.com/netsp/the-dangers-and-promise-of-multipath-tcp.html.

[12] Bagnulo, Marcelo. "RFC 6181: Threat Analysis for TCP Extensions for Multipath Operation with Multiple Addresses." Internet Engineering Task Force (2011).

[13] Ford, Alan, Costin Raiciu, Mark Handley, Sebastien Barre, and Janardhan Iyengar. "Architectural guidelines for multipath TCP development." IETF, Informational RFC 6182 (2011): 2070-1721.

[14] Fall, Kevin R., and W. Richard Stevens. TCP/IP illustrated, volume 1: The protocols. addison-Wesley, 2011.

[15] Singh, Ashutosh, Carmelita Goerg, Andreas Timm-Giel, Michael Scharf, and T-R. Banniza. "Performance comparison of scheduling algorithms for multipath transfer." In Global Communications Conference (GLOBECOM), 2012 IEEE, pp. 2653-2658. IEEE, 2012.

[16] Raiciu, Costin, Christoph Paasch, Sebastien Barre, Alan Ford, Michio Honda, Fabien Duchene, Olivier Bonaventure, and Mark Handley. "How hard can it be? designing and implementing a deployable multipath TCP." InProceedings of the 9th USENIX conference on Networked Systems Design and Implementation, pp. 29-29. USENIX Association, 2012.

[17] Eardley, Philip. "Survey of MPTCP implementations." (2013).

[18] Blanchet, Marc, and Pierrick Seite. "Multiple interfaces and provisioning domains problem statement." (2011).

[19] Bagnulo, Marcelo, Olivier Bonaventure, Fernando Gont, Christoph Paasch, and Costin Raiciu. "Analysis of MPTCP residual threats and possible fixes."Analysis (2013).

[20] Barré, Sébastien, Christoph Paasch, and Olivier Bonaventure. "Multipath TCP: from theory to practice." In NETWORKING 2011, pp. 444-457. Springer Berlin Heidelberg, 2011.

AUTHORS

Adwan Yasin is an associate Professor, Former dean of Faculty of Engineering and Information Technology of the Arab American University of Jenin, Palestine. Previously he worked at Philadelphia and Zarka Private University, Jordan. He received his PhD degree from the National Technical University of Ukraine in 1996. His research interests include Computer Networks, Computer Architecture, Cryptography and Networks Security.

Hamzah M. A. Hijawi received his BS in computer system engineering from Birzeit University, Ramallah, Palestine, in 2011. He is currently working with Exalt technologies as a software engineer since 2010, and currently pursuing his Master of Computer Science from Arab American University, Jenin, Palestine. His researches interest include Computer Networks, Information Security, Wireless Sensor Networks and Data Mining.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 6, June 2016

TEMPORAL PERFORMANCES EVALUATION OF MULTI-ROBOT DEMINING SYSTEM INSPIRED BY ANT BEHAVIOR.

Riadh SAAIDIA, Mohamed Sahbi BELLAMINE, and Abdessattar BEN AMOR

Abstract— **In this paper we adopt a cooperative strategy based on ACO (Ant Colony Optimization) algorithms to coordinate a Multi Robots System (MRS). Our principal objective is to evaluate temporal performances for this system by choosing demining operations as a benchmark problem. In this work, we try to adapt the ACO algorithm parameters for different mine distribution in order to reduce time demining operations. In particular, we report effects of evaporation pheromone rate model and minefield configuration on temporal performances.**

***Index Terms*— ACO algorithms, multi-robot system (MRS), evaporation pheromone rate, demining system.**

I. Introduction

As stated in [1], the percentage of human victims and deaths caused by mine, improvised explosive device (IED) and explosive remain of war (ERW) has been declining since 1999. However, mine accidents number is still important, especially if we compare the civilian causalities percentage with military one, we find that it has risen for 73% in 2011 to 78% in 2012.

In 2012, the landmine report witnessed a high total number of 3628 mine/ERW/IED casualties especially among children and women. Also there is a detection of 1066 killed people and 2552 injuries. Despite all these figures, the real number of casualties is still unknown and related to world struggle. Although the clearness of landmine represents a recurrent problem because, the undamaged surface is extended yearly, and it needs efficient methods to ensure the clearance goal.

At least, both the standard demining clearance model operations (UNDHA) and Mine Action Standards (IMAS) must ensure 99.6% and 100% of successful mine detection [2-4].

Taking into consideration, the importance of personal safety even before timing demining process performances, the robots is used to replace the manual methods, in order to save the human being and improve the activity by speeding up reliably and safely the demining process.

In order to achieve these goals, it must pay attention to the nature of landmine and the characterization of demining instruments, also it must use different types of sensors and equipment of detecting landmines. The application of robotic research to demining operations purposes requires the integration of various technologies, including demining-oriented functions like the adaptability to field mines distributions, type of control architecture, integration of heterogeneous sensors, autonomous navigation , coordination in the case of multi-robots system, communication implementation, Machine intelligence and signal processing algorithms [2].

The operation of exploring unknown configuration minefield faces some difficulties which are: the limited performances of the existing robotic systems [5], also the highly sophisticated technology instrument on the robots [6].

In addition, timing optimization in this operation presents a challenge that must be taken into account because of its relation to humanitarian objective [7]. So in order to ensure the security restrictions different assistant devices were added to the goal of limiting the risk of human error and rising the estimation of risk zone. However, the objective is still hard to be fullfiled because of the sophisticated robot agents and the mines distribution variety which enhance the demining operations cost.

In this paper, there is a presentation of different applications of multi-robot systems, which are adapted to minimize the time detection of mines proportion (Mx%=90%) [8]. Due to the importance and complexity of the demining operations, it is obvious and necessary to adapt an efficient coordination algorithm. So, in this work, we adopt Ant colony optimization

Paper submission date: 05/25/2016. This **work** was supported in part by the Computer Laboratory for Industrial Systems (LISI), National Institute of Applied Sciences and Technology (University of Carthage).

Riadh SAAIDIA is with the Computer Laboratory for Industrial Systems (LISI), National Institute of Applied Sciences and Technology (University of Carthage), INSAT, centre urbain nord BP 676 Tunis, Tunisia 1080, TN TUNISIA (e-mail: Riadh.saaidia@insat.rnu.tn).

Mohamed Sahbi BELLAMINE is with the Computer Laboratory for Industrial Systems (LISI), National Institute of Applied Sciences and Technology (University of Carthage), INSAT, centre urbain nord BP 676 Tunis, Tunisia 1080, TN TUNISIA (e-mail: aroussia@insat.rnu.tn).

Abdessattar BEN AMOR is with the Computer Laboratory for Industrial Systems (LISI), National Institute of Applied Sciences and Technology (University of Carthage), INSAT, centre urbain nord BP 676 Tunis, Tunisia 1080, TN TUNISIA (e-mail: mrdssrst@yahoo.fr)

algorithm as an example of coordination algorithm based on meta-heuristic algorithms to treat complexity of demining problem and scale of landmine fields [9, 10].

The remainder of this paper is organized as follows. Sect. 2 focuses on different works where multi-robots are applied to ensure demining operations. In the case of mine distribution, type of meta-heuristics used for collaboration algorithms and performances metrics. Sect. 3 presents the field mine distribution and collaboration models used in demining operations. Sect. 4 describes the simulation considerations for performed experiences. Sect. 5 lists and analyzes the simulations results. Sect. 6 is reserved for results discussion.

II. Related works

Multi-robots application in demining operations for humanitarian purposes represents an evaluation example of coordination strategy performance. Many researches such as [11-13] use specific coordination strategy in order to evaluate some criteria performances. General research organization starts with the definition of collaboration algorithms used in order to perform specific task. Demining process, which is highlighted in this research, includes many constraints related to the nature of minefield distribution and performance evaluation criteria. Some researches as in [11, 13, 14] give statistical studies on variety of spatial mine distribution in minefield. In fact, mines field spatial distributions in conflict zones are highly complex and varied. Landmine descriptions cannot be defined easily with deterministic clustering approaches. Landmine variety induces different mine distribution patterns, that one can be used to test hypotheses for demining operations. However, other assumptions have influence on performances evaluation systems. Combining the different parameters (incidents, populations, roads, agriculture field, etc.) for defining minefield map, would allow the consideration of environmental and social conditions [7].

Simulation example given in [5] tests real case minefield distributions in order to realize an automatic estimator to mines localization. Mines distribution configuration represents a limitation in the case of unknown mined environment. Nevertheless, in several cases, mines distribution can be modeled by stochastic model like in [6, 7, 14]. Moreover, the efficiency of demining operations depends on the scenario followed for each robotic agent.

On the other hand, the choice of collaboration strategy represents other constraints. In fact, demining operations with multi-robots systems raise complexity of collaboration interactions [11, 15]. In this case, the application of suitable meta-heuristic algorithms for multi-robot demining operations was performed in research such as [16-19]. Research studies focus on combined and modified heuristic (as is the case for Genetic algorithms, ACO algorithms, etc.) to enhance general performances of multi-robots systems.

As a result, studies as [20] define some evaluation metrics to quantify collaboration performance cost. Localization and distribution robotic agents configuration were taken as evaluation criteria. These criteria depend on the application of constraints like possible robot agents interference [21]. A set of generic performance metrics was employed to evaluate each aspect of robotic demining systems. These performance metrics include demining processing speed to measure time elapsed until demining operations can be totally or partially achieved. The rest of experimentations focus on temporal performance optimization by using modified meta-heuristic algorithms.

In particular, configuration parameters for minefield and coordination algorithm heuristic, as type of mine distributions and effects of evaporation pheromone rate, were treated in experimentations. Other performance metrics like: robotic agents displacements which represents aggregation of the distances inter-agent position during the demining operations (consumed energy), robotic Agents proportion of agents which ensure demining operations, robotic group size effect and communication flow exchanged between agents during robots interactions; represent other optimization objectives and they will be treated in further works.

III. Methods and hypothesis

This part represents general configuration parameters for tested environment. These parameters include minefield distribution and adaptation of ACO algorithms for collaborative demining robotic foraging. The measurement of demining operations time was performed at different values of configuration parameters. Tested mines proportion (Mx %) has been fixed to 90% for a total number of 50 mines [6].

At the first level, robots/mines ratio (RM%) is tested as an influential parameter for time system performances. At the second level, different configurations of minefield distribution were evaluated. At the third level, evaporation pheromone model is studied as influential parameter for research navigation model based on ACO algorithms [22]. The evaporation pheromone rate is increased gradually and the operation of detection mines time is noted.

A. Mine configuration

The mine spatial distribution has possible effect in mine detection time [6, 7]. The performance of different collaborative navigation methods is evaluated by the consideration of three types of distribution models. These distributions include random distribution, fixed spatial distribution and random line distribution.

In the case of random distribution mines are placed randomly with uniform density of probability [23, 24], the second type of distributions are reserved to fixed mine position [25]. Two different dispositions with limited mined zone are evaluated. That type of distribution is based on normal mixture model (Figure. 2 and 3)[26, 27]. The definition of the mined zones in the version of fixed distributions depends on matrix variance normal distribution (Fixed 1: $\sigma_1^2=1$ and $\sigma_2^2=16$;Fixed 2: $\sigma_1^2=10$ and $\sigma_2^2=16$)[28].

As presented in [29], and in the case of environment symmetry the localization represents a complicated task. This complexity is due to the correctness of robot position and orientation estimation (unknown mine land without specific information). Collaborative algorithms, as for ACO

algorithms, can reduce elapsed time in mines research operations.

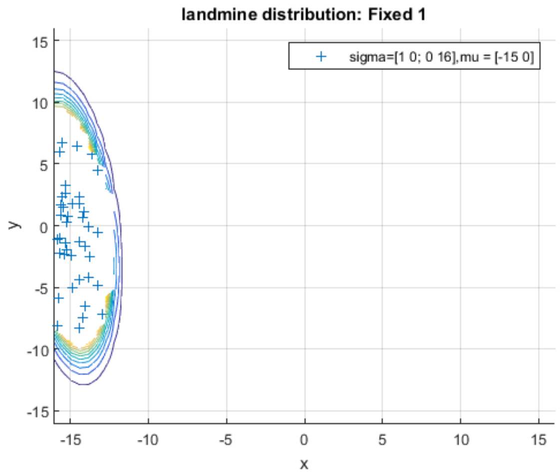


Fig. 1. Fixed spatial distribution 1.

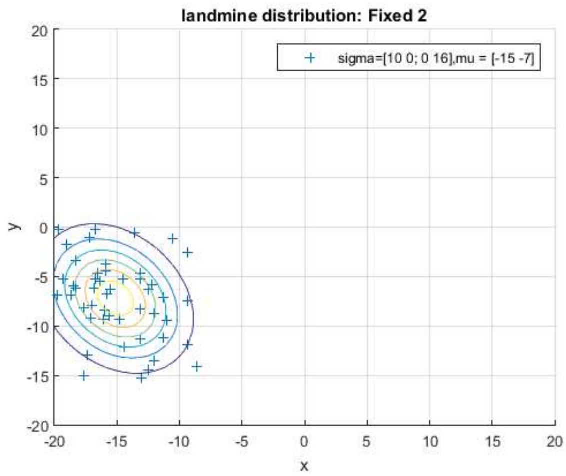


Fig. 2. Fixed spatial distribution 2.

In the case of random line distribution, mine lines are randomly placed along the line or dropped with a constant spacing. The random lines are given a very broad margin of placement error. The random spacing lines are assumed to represent positioning errors mainly due to navigation and drop timing errors. This distribution is based on Poisson mixture model [30, 31] with the probability to find a mine at the x position on the projected line is expressed as follow [5]:

$P(X_N <= x)=(1- e^{-\lambda x})^N$ (1)

With N is the number of mine detected and λ is the Poisson rate.

Random lines are assumed to have random orientation and mine spacing. But in these experimentations; random mine lines are parallel [5].

B. Navigation and research methods

This part includes the presentation of mine research methods adopted by different robot agents. The evaluation of this methods effect is based on the time detection mines quality. In this experimentation, three main collaborative navigation algorithms were performed including random research model (BASE), ant research model (AS-ACO) and modified ant research model (M-AS-ACO).

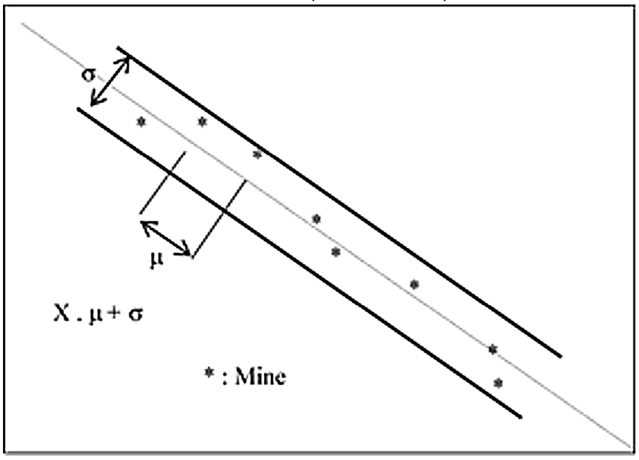


Fig. 3. Random line distribution (s=1,μ=3 and areas dimensions=16x16).

In the case of the BASE model, robot agents do not adopt a particular logic for mine research. So robot agents are not restricted to any constraint except some particular rules listed as fallows:

– R1: when a robot agent finds a mine, it must return to the base for the deactivation of mine operation.
– R2: used base is fixed.
– R3: all robot agents are placed in the base at the demining operations beginning.

The robot agents of the AS-ACO model adopt a mine research strategy based on ACO algorithm to find optimum demining operation. The same rules adopted in BASE model (R1, R2 and R3) are retained. The used robot agents' path is fixed by pheromone rate τ deposited by other searching agents. Three main methods are adopted for pheromone rate calculation:

1) 1st case:

In this test, the evaporation pheromone rate ρ (static evaporation pheromone rate) is fixed and the pheromone rate calculation is given as follows [32]:

$\tau(k)=\tau(k-1)(1-\rho)$ (2)

2) 2nd case:

This ACO algorithm configuration adopts a programmable evaporation pheromone rate (dynamic evaporation pheromone rate) to calculate pheromone rate as follows:

$\tau(k))=\tau(k-1)(1-\rho)+(1-(1+Q)^{-1}) \tau(k-1)$ (3)

$\rho=(1+(\tau-\alpha)^4.(2\alpha)^{-0.5})$, where $\alpha=0.5$ (4)

Equation (3) introduces a heuristic Q factor, which represents an algorithm quality factor [22]. The α factor used in programmable evaporation pheromone rate was fixed to 0.3. The Q appreciation factor for method research rule is formulated as follows [11]:

$Q=TP.(TP+FN)^{-1}.TN.(FP+TN)^{-1}$ (5)

Equation (5) introduces two main rules for demining research operations:

- Dynamic rule 1= mine research operation (TP=find mine when trying to research mine, FP = robot does not find mine when trying to research mine)
- Dynamic rule 2= base return (TN = robot already

charging mine in return when trying to return to base, FN = mine discharged into the base)

3) *3rd case:*

The navigation model in this case adopts also a programmable evaporation pheromone rate (timed evaporation pheromone rate). But, the evaporation pheromone rate is defined by the determination of wasted time elapsed between two successive mine detections as follows:

$$\rho = (1+tM1)^{-1}.\Delta t \quad (5)$$

$$\Delta t = tM1-tM2+1 \quad (6)$$

Where tM1=detection time for $mine_i$ and tM2= detection time for $mine_{i-1}$

The method adopted by M-AS-ACO model is also based on the ACO algorithm. This model considers a mobile base in order to minimize base-mine displacement. Base coordinates are defined by P_x and P_y:

$$P_x(k)=0.5\ (P_x(k-1)+R_{ix}(k)) \quad (7)$$

$$P_y(k)=0.5\ (P_y(k-1)+R_{iy}(k)) \quad (8)$$

The $(R_{ix}(k), R_{iy}(k))$ couple represents the coordinates of recent detected $mine_i$. The idea presented was inspired by the intensification and diversification [33, 34]. The diversification for robotic agent represents the ability to demine many and different mine land regions. Intensification is summarized in the ability of base guides demining operation in specific zones with high mine concentration. At this stage, the robot agents are reserved for mine research and the deactivating operations are assigned to the base as a new agent type.

IV. Simulation Protocol

This section introduces general simulation protocols followed in collaborative algorithms efficiency validation. All simulations are performed with NetLogo [35, 36]. NetLogo is used as a software platform to simulate robotic agents and landmine map. In fact, NetLogo supports advanced modeling of complex systems using a library of java programming primitives. In NetLogo simulation environment, robotic agents are modeled in simple design without the consideration of collision avoidance.

As given in Table 1: the experience design was performed by variation of the evaporation pheromone rate and kind of landmine distributions. Each experience is repeated ten times using NetLogo API control. The mine detection time values was reported to MATLAB software platform in order to compare different configuration results.

A simplified foraging scenario was taken to describe demining operations. Robots states include the searching and homing state. When a robot detects a mine, it picks it up and comes back toward neutralizing base. Execution demining time is accounted while a robot is either in searching or homing mode. Time of other robots avoidance is not considered in demining scenario. Fig. 4 shows the state diagram for demining operations scenario. Robotic agents detect, collect mines and bring them to a mine neutralizing base.

V. Result

Experimental studies in this manuscript were performed for different RM% ratio. According to [21], rising RM% ratio beyond some limits do not affect time detection because of the interference of robotic agents, which stabilizes the time result. In order to test evaporation pheromone rate influence on time demining optimization; some tests are performed with different RM% ratio. These tests identify limits that do not modify temporal performances. Additional experimentations, that perform the application of various RM% rate on presented mines distributions and collaboration models based on ACO algorithms, were conduct to verify the hypotheses. The rising robotic agents number (in order to minimize mine detection time) has no influence on system timing performances. Fig. 5 gives an example of time detection mine stabilization for BASE model with different distributions and RM%.

TABLE I
SIMULATION PARAMETERS

Model	Evaporation pheromone rate %	Distributions
AS-ACO	0%-100%	Random, fixed 1, fixed 2 and random line
M-AS-ACO	0%-100%	Random, fixed 1, fixed 2 and random line

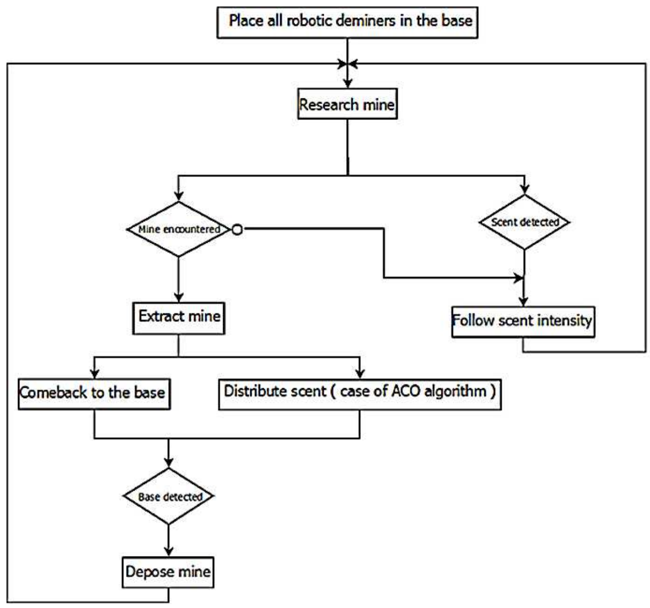
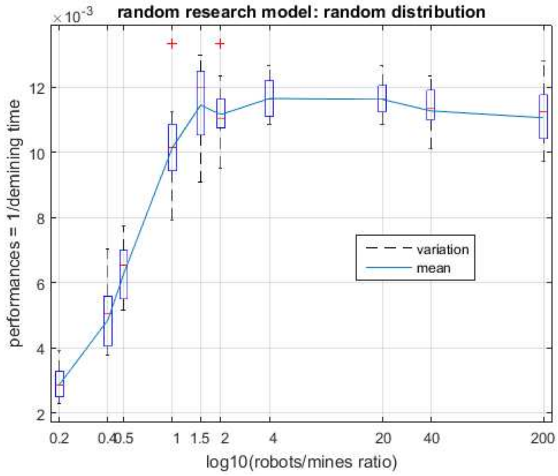


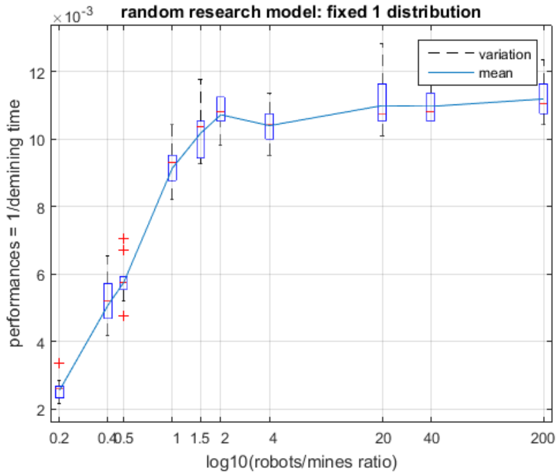
Fig. 4. Behavior diagram of a multi-robot demining system.

This part presents the possible effect of evaporation pheromone rate variation on demining time performances for both AS-ACO and M-AS-ACO algorithms (Mx%=90%). In each experimentation, pheromone evaporation rate is increased regularly by 10%.

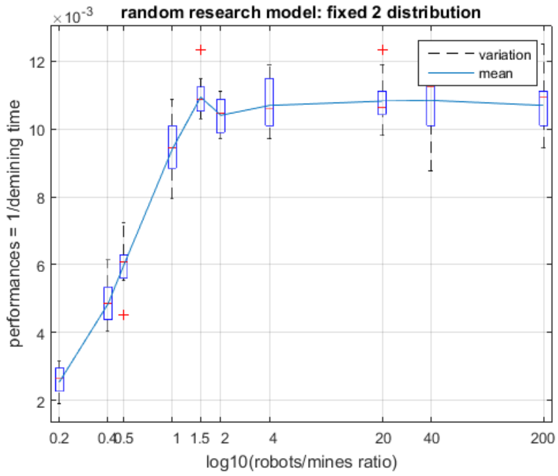
https://sites.google.com/site/ijcsis/
ISSN 1947-5500



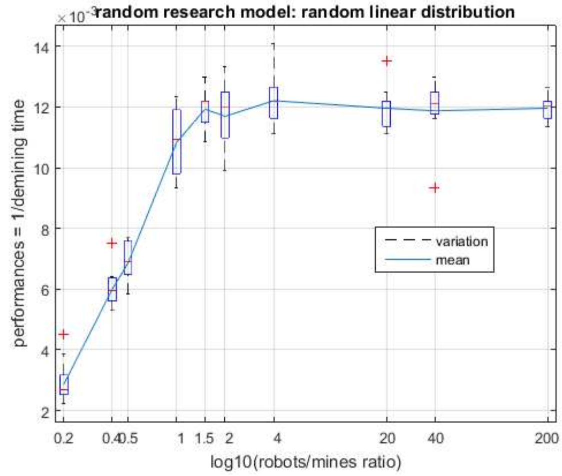
(a). random distribution.



(b). Fixed 1 spatial distribution.



(c). Fixed 2 spatial distribution.



(d). random line distribution.

Fig.5: Demining MRS performances (1/demining time) for different mine distributions in the case of BASE model

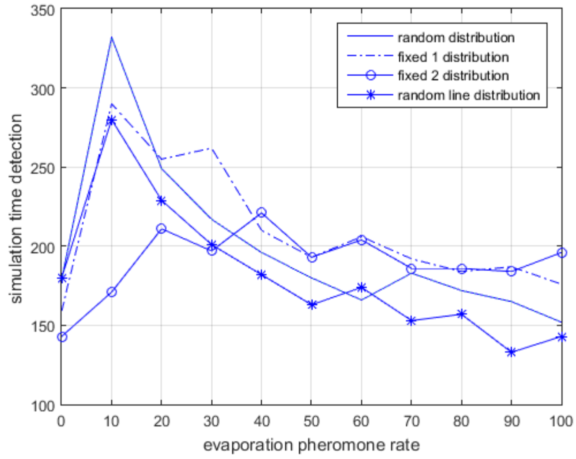


Fig. 6. Time detection results for the AS-ACO model

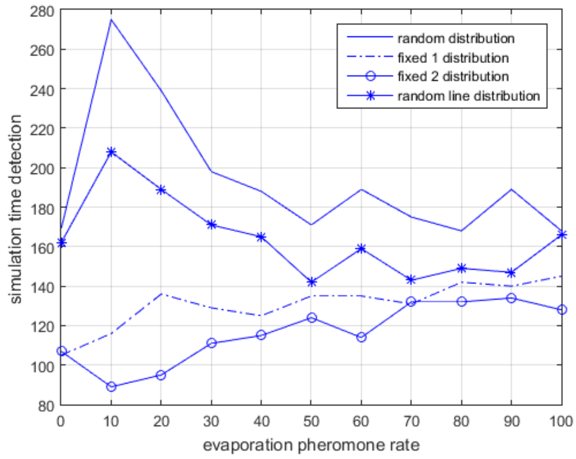


Fig. 7. Time detection results for the M-AS-ACO model

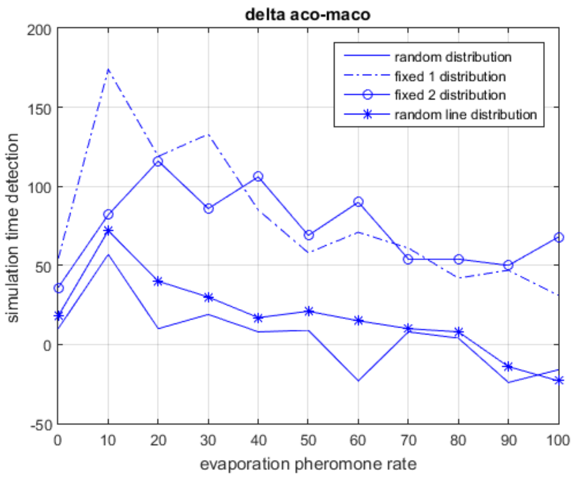


Fig. 8. Time detection comparison between AS-ACO and M-AS-ACO models

Fig.6 and 7 represent the detection time variation relating to the minefield distribution type for both AS-ACO and M-AS-ACO models. For lower pheromone evaporation rate, higher values of detection time results are taken with random distribution. The rising pheromone evaporation rate ameliorates temporal performances. However, this decrease of mine detection-time is stabilized for high evaporation. In fact, detection time results are limited to a range of 200 s.t for evaporation pheromone rate > 60% in the case of AS-ACO model and for evaporation pheromone rate > 30% in the case of M-AS-ACO model.

Fig. 8 indicates the time variation between AS-ACO and M-AS-ACO models. Considering the effect of minefield distribution type separately, M-AS-ACO model presents better timing results than AS-ACO model with lower pheromone evaporation rate. AS-ACO model presents better timing results than M-AS-ACO model only in the case of fixed spatial distributions with high pheromone evaporation rate (>80%).

The impact of pheromone evaporation rate on time system performances is noted at the beginning of the solutions construction. Adopting a programmable pheromone evaporation rate which induces new solution explorations should reduce time demining. Researches of [22, 37, 38], use different models of programmable evaporation rate based on a mathematical formulation. Dealing with the evaporation pheromone example given by [22], this model is taken as a reference to evaluate our evaporation pheromone rate model. Simplifying evaporation pheromone model is the principal motivation of selection of a timed algorithm model.

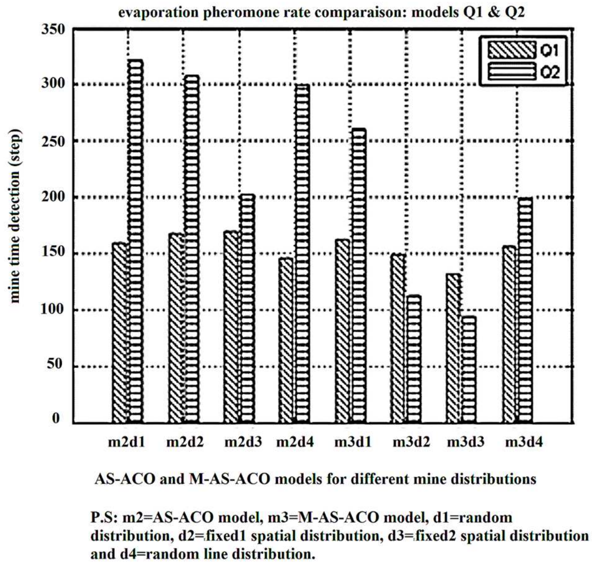


Fig. 9. Evaporation pheromone rate model comparison

Fig. 9 reports the temporal result difference between different evaporation pheromone models for AS-ACO and M-AS-ACO collaborative algorithms. Mathematical evaporation pheromone rate model [22] is represented by Q1 model. Our evaporation pheromone rate model is represented by Q2 model. In the case of AS-ACO model (m2d1, m2d2, m2d3 and m2d4); temporal results obtained with Q1 model are better than with Q2 model except the result in fixed 2 distribution (m2d1). In fact, the system equipped with Q2 evaporation pheromone model takes double time to detect 90% of mines compared to Q1 model. This different change in the case of M-AS-ACO model and better temporal performances is detected with Q2 model in the case of fixed distributions. Multi-robot system experimentations are performed on the software simulation platform. In real implementation, the application of mathematical complex model for evaporation pheromone rate should require more hardware resources and reduce temporal performances.

VI. Discussion

The realized experimentations use a fixed setting of RM% rate. Generally, rising RM% rate is higher than 50% does not enhance cooperation impact on demining time optimization. These results were treated also in the previous researches [21].

The principal aim of research in this paper is the connection between evaporation pheromone rate and timing performance. In fact, as given in Fig. 6, 7 and 8 better timing results are detected for M-AS-ACO model (in most studied cases: Table 2).

TABLE 2: SUMMARY OF TIME RESULT VARIATION BETWEEN AS-ACO AND M-AS-ACO MODELS

Distribution	0%-50%	50%-70%	70%-80%	80%-100%
Random	+	-	+	-
Fixed 1	+	+	+	+
Fixed 2	+	+	+	+
Random line	+	+	-	-

(+/-) Sign of time result variation between AS-ACO and M-AS-ACO models for different static evaporation pheromone rates ($time_{AS\text{-}ACO} - time_{M\text{-}AS\text{-}ACO}$)

In general, ACO algorithms are made from ant foraging behavior. ACO optimization gives a short path solution to one source of food. In the case of demining problems, the mines are distributed in various positions. The best initial situation ACO algorithm consists of a limited zone mine concentration. This situation is given by fixed1 and fixed2 distributions. For these two mine distributions and at a lower evaporation pheromone rate, better timing results are obtained in comparison to the base model. However, with random distributions (random and random line distributions), time demining results are degraded with AS-ACO model in favor of the BASE or M-AS-ACO model. The Amelioration of the AS-ACO model results is given by the raising evaporation pheromone rate. In fact, this action helps robotic agents to forget the previous detected mine positions and forces the agents to explore new zones. Time result experimentations are reduced for the evaporation pheromone rate, which are higher than 60% in the case of AS-ACO model, and 30% rate in the case of M-AS-ACO model. The solution is ensured by M-AS-ACO model presents flexibility toward different mine distributions.

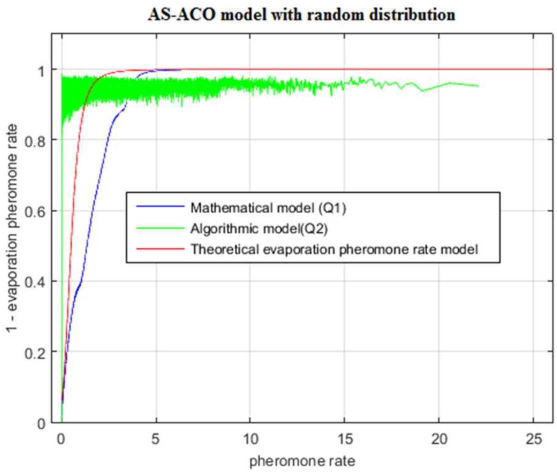
The variation of the evaporation pheromone rate has an impact on timing results. With this interpretation, some researchers [22, 39] applied a specific function to define the evaporation pheromone rate. In general, this function is bounded between 0 and 1. It rises exponentially with the pheromone rate. Our proposed evaporation pheromone rate Q2 gives lower timing performances for demining operations in the case of the AS-ACO model. The worst timing results are detected for random mine distribution (55% of time result reduction). However, the Q2 model gives better timing results in the case of the M-AS-ACO model with fixed mine distributions. The best results are detected for fixed 2 mine distribution. The evaporation pheromone Q1 model still has better results in random distributions (with M-AS-ACO model) but the timing performance differences between Q1 and Q2 models are reduced in comparison to AS-ACO model.

TABLE 3: COMPARISON TIME RESULT BETWEEN Q1 AND Q2 MODELS

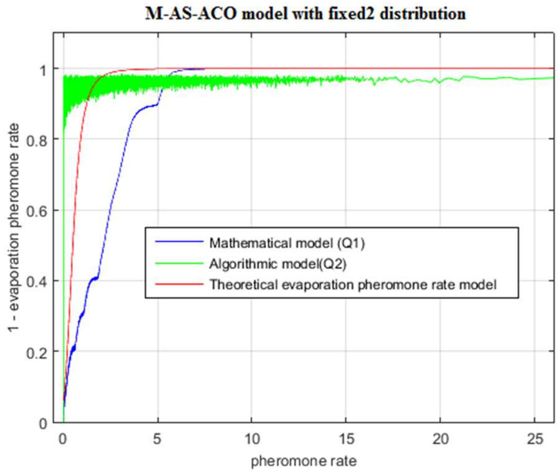
Distribution	AS-ACO model	M-AS-ACO model
Random	55%	32%
Fixed 1	46%	-8%
Fixed 2	12%	-27%
Random line	42%	28%

(*) %=($time_{Q2}$-$time_{Q1}$)/ $time_{Q2}$

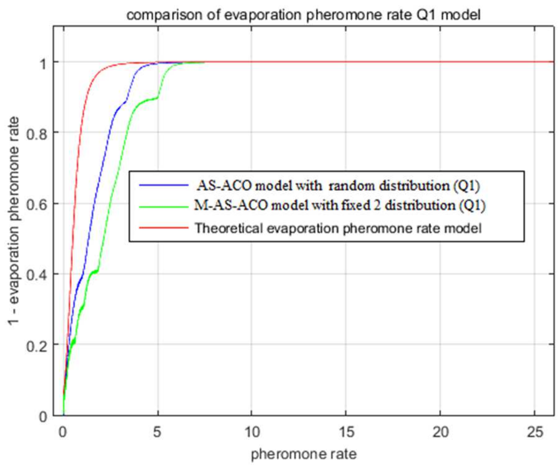
To explain the results given by Table 3, the worst and the best result for Q2 model are selected. The worst time result corresponds to the AS-ACO cooperative model with random distribution. The best time result corresponds to the M-AS-ACO cooperative model associated with fixed 2 mine distribution.



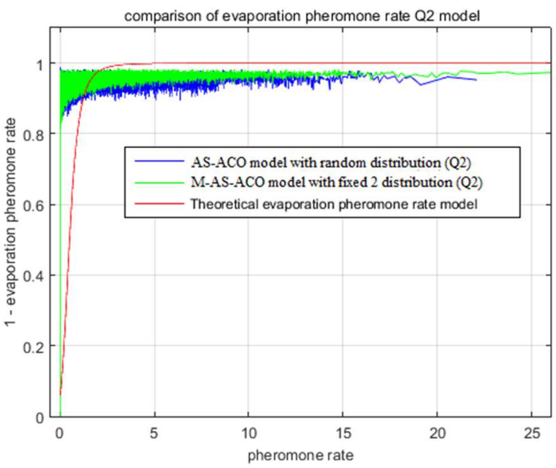
(a) AS-ACO model with random distribution



(b) M-AS-ACO model with fixed 2 distribution



(c) Comparison of evaporation pheromone rate Q1 model



(d) Comparison of evaporation pheromone rate Q2 model

Fig. 10. Evaluation of the evaporation pheromone rate model (Q1 and Q2 models) for AS-ACO and M-AS-ACO model

Fig. 10 reports the variation of the evaporation pheromone rate models in the worst time result (Fig. 10.a) and the best time result (Fig. 10.b). The recorded evaporation pheromone rate from Q1 model simulations differs from theoretical evaporation pheromone rate formulation (4). This difference is amplified for the M-AS-ACO model. In addition, the model guided by Q2 approaches the theoretical model but it presents higher sensitivity of the pheromone rate variation and saturates fast bounded limit. Fig. 10.c gives a comparison between Q1 model in the AS-ACO and M-AS-ACO model. Evaporation pheromone model converges to the theoretical model with additional delay in the M-AS-ACO model. In Fig. 10.d, the Q2 model preserves the same pattern and therefore gives better time results for fixed distributions.

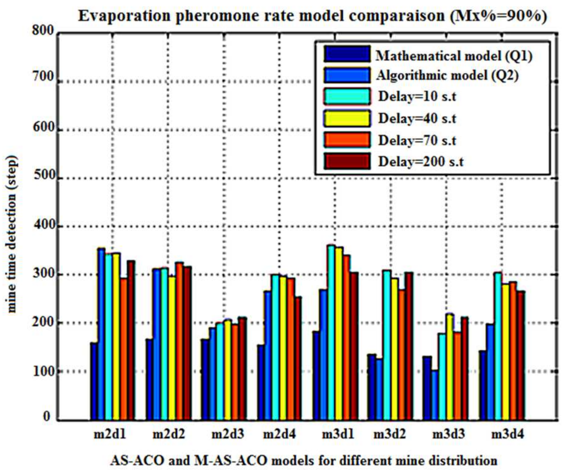


Fig. 11. Time results for different models of evaporation pheromone rate

Fig. 11 presents the time demining results for the reduction of evaporation pheromone rate sensitivity to variation of the pheromone rate. These attempts of Q2 model amelioration are based on the introduction of delay in the iterations of evaporation pheromone rate calculation. Some increasing values of delays (10 s.t, 40 s.t, 70 s.t and 200 s.t) are experimented. The general time performances of the demining system is degraded for the AS-ACO and M-AS-ACO models and there is no modification of evaporation pheromone rate pattern in the function of pheromone rate.

VII. Conclusion

This paper presents the experimentations of the pheromone evaporation rate on the multi-robotic demining system. The effects of the pheromone evaporation rate are noted for particular rates and better results are obtained with M-AS-ACO algorithms. The temporal performance of demining multi-robot systems is obtained by modifying the ACO algorithms. However, results are still depending on the environment configurations and on the other modifications can be performed on ACO algorithms especially by studying the pheromone evaporation rate.

The application of programmable evaporation pheromone rate helps to improve temporal performances. The improvement of temporal performances is set up with the evaporation pheromone rate pulse (instead of high evaporation pheromone rate maintain). The choice of the model of evaporation pheromone rate modifies temporal performances of the demining system. The proposed evaporation pheromone rate Q2 enhances temporal performances of the demining operations for a particular configuration mainly with the M-AS-ACO model and fixed mine distribution. The studied Q1 model is an example of programmable evaporation pheromone rate. Other functional models can be tested. The aim of the algorithmic evaporation pheromone model is to simplify the implementation of this system. In our case, the additional experimentations on real implementation of multi-robot controller must be performed to evaluate the algorithmic model of evaporation pheromone rate. A collaborative model based on Ant Colony Optimization is selected. In addition, other meta-heuristic algorithms can be applied in the same case. In particular, hybrid meta-heuristic algorithms should be experimented on multi-robotic controllers.

References

[1] ICBL-CMC, "landmine monitor report," International Campaign to Ban Landmines – Cluster Munition Coalition (ICBL-CMC)2015.

[2] M. K. Habib, "Humanitarian demining: Reality and the challenge of technology-the state of the arts," *International Journal of Advanced Robotic Systems*, vol. 4, pp. 151-172, 2007.

[3] J. F. V. Silva, S. Jaramillo, L. J. Ramirez, C. O. Valderrama, and C. E. Restrepo-Garces, "Perspectives to pain associated to land mine injury," *Techniques in Regional Anesthesia and Pain Management*, vol. 16, pp. 175-179, 2012.

[4] N. E. Walsh and W. S. Walsh, "Rehabilitation of landmine victims: the ultimate challenge," *Bulletin of the World Health Organization*, vol. 81, pp. 665-670, 2003.

[5] J. Hyland and C. Smith, "Nonparametric spatial analysis for mine line detection," in *OCEANS 2009, MTS/IEEE Biloxi-Marine Technology for Our Future: Global and Local Challenges*, 2009, pp. 1-10.

[6] C. A. G. Díaz and A. P. Quimbaya, "Spatial processing of events with landmines in Colombia," in *7th Colombian Computing Congress (CCC)* 2012, pp. 1-6.

[7] A. C. Alegría, H. Sahli, and E. Zimányi, "Application of density analysis for landmine risk mapping," in *IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM)*, 2011, pp. 223-228.

[8] A. R. Thompson, "Evaluating the Combined UUV Efforts in a Large-Scale Mine Warfare Environment," DTIC Document2015.

[9] G. D. Tipaldi and K. O. Arras, "Planning Problems for Social Robots," in *ICAPS*, 2011.
[10] V. Munirajan, "A swarm intelligence based approach to the mine detection problem," 2004.
[11] A. ElGindy and A. Khamis, "Team-theoretic approach to cooperative multirobot systems: Humanitarian demining as a case study," in *International Conference on Engineering and Technology (ICET)*, 2012, pp. 1-6.
[12] M. K. Habib, "Development of robot and navigation techniques for humanitarian demining," in *6th IEEE International Conference on Industrial Informatics, 2008. INDIN 2008.* , 2008, pp. 418-423.
[13] M. A. Jaradat, M. N. BaniSalim, and F. H. Awad, "Autonomous navigation robot for landmine detection applications," in *Mechatronics and its Applications (ISMA), 2012 8th International Symposium on*, 2012, pp. 1-5.
[14] C. Smaili, M. E. B. El Najjar, and F. Charpillet, "A hybrid bayesian framework for map matching: formulation using switching Kalman filter," *Journal of Intelligent & Robotic Systems,* vol. 74, pp. 725-743, 2014.
[15] Y. J. Stergiopoulos and A. Tzes, "Decentralized swarm coordination: A combined coverage/connectivity approach," *Journal of Intelligent & Robotic Systems,* vol. 64, pp. 603-623, 2011.
[16] X. Chen, Y. Kong, X. Fang, and Q. Wu, "A fast two-stage ACO algorithm for robotic path planning," *Neural Computing and Applications,* vol. 22, pp. 313-319, 2013.
[17] J. Lu, N. Wang, J. Chen, and F. Su, "Cooperative path planning for multiple UCAVs using an AIS-ACO hybrid approach," in *Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology*, 2011.
[18] K. M. Salama and A. M. Abdelbar, "Exploring different rule quality evaluation functions in aco-based classification algorithms," in *Swarm Intelligence (SIS), 2011 IEEE Symposium on*, 2011, pp. 1-8.
[19] B. Shuang, J. Chen, and Z. Li, "Study on hybrid PS-ACO algorithm," *Applied Intelligence,* vol. 34, pp. 64-73, 2011.
[20] N. Bredeche and Y. Chevaleyre, "The robot swarm re-localization problem," in *Robotics and Biomimetics, 2008. ROBIO 2008. IEEE International Conference on*, 2009, pp. 1876-1881.
[21] G. Pini, A. Brutschy, M. Birattari, and M. Dorigo, "Task partitioning in swarms of robots: reducing performance losses due to interference at shared resources," in *Informatics in Control Automation and Robotics*, ed: Springer, 2011, pp. 217-228.
[22] P. Lalbakhsh, M. Fasaei, and M. N. Fesharaki, "Focusing on rule quality and pheromone evaporation to improve ACO rule mining," in *Computers & Informatics (ISCI), 2011 IEEE Symposium on*, 2011, pp. 108-112.
[23] J. M. M. Anderson, "A Generalized Likelihood Ratio Test for Detecting Land Mines Using Multispectral Images," *IEEE Geoscience and Remote Sensing Letters,* vol. 5, pp. 547-551, 2008.
[24] W. Jing, L. Xiaoling, and N. Jianhong, "Probability density function estimation based on representative data samples," in *Communication Technology and Application (ICCTA 2011), IET International Conference on*, 2011, pp. 694-698.
[25] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, 2004, pp. 28-31 Vol.2.
[26] C. Schultz, A. C. Alegría, J. Cornelis, and H. Sahli, "Comparison of spatial and aspatial logistic regression models for landmine risk mapping," *Applied Geography,* vol. 66, pp. 52-63, 1// 2016.
[27] A. R. Gonçalves, F. J. Von Zuben, and A. Banerjee, "Multi-task Sparse Structure Learning with Gaussian Copula Models," *Journal of Machine Learning Research,* vol. 17, pp. 1-30, 2016.
[28] C. B. Do, "The multivariate Gaussian distribution," 2008.
[29] F. Abrate, B. Bona, M. Indri, S. Rosa, and F. Tibaldi, "Multirobot localization in highly symmetrical environments," *Journal of Intelligent & Robotic Systems,* vol. 71, pp. 403-421, 2013.
[30] L. H. Moulton and A. A. Benini, "Community level risk factors for numbers of landmine victims in Chad and Thailand," *Journal of Epidemiology and Community Health,* vol. 57, pp. 956-959, December 1, 2003 2003.
[31] M. Nagode, "rebmix: Finite Mixture Modeling, Clustering & Classification," 2016.
[32] O. Castillo, R. Martínez-Marroquín, and P. Melin, "Bio-inspired optimization of fuzzy logic controllers for robotic autonomous systems with PSO and ACO," *Fuzzy Information and Engineering,* vol. 2, pp. 119-143, 2010.
[33] M. Lozano and C. García-Martínez, "Hybrid metaheuristics with evolutionary algorithms specializing in intensification and diversification: Overview and progress report," *Computers & Operations Research,* vol. 37, pp. 481-497, 2010.
[34] J. Jin, T. G. Crainic, and A. Løkketangen, "A cooperative parallel metaheuristic for the capacitated vehicle routing problem," *Computers & Operations Research,* vol. 44, pp. 33-41, 2014.
[35] A. Hussain and M. Niazi, "Toward a formal, visual framework of emergent cognitive development of scholars," *Cognitive Computation,* vol. 6, pp. 113-124, 2014.
[36] S. Tisue and U. Wilensky, "NetLogo: Design and implementation of a multi-agent modeling environment," in *Proceedings of agent*, 2004, pp. 7-9.
[37] O. Gokalp and A. Ugur, "Improving performance of ACO algorithms using crossover mechanism based on mean of pheromone tables," in *Innovations in Intelligent Systems and Applications (INISTA), 2012 International Symposium on*, 2012, pp. 1-4.
[38] H. Huang, C.-G. Wu, and Z.-F. Hao, "A pheromone-rate-based analysis on the convergence time of ACO algorithm," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 39, pp. 910-923, 2009.
[39] C.-T. Cheng, K. Fallahi, H. Leung, and C. K. Tse, "Cooperative path planner for UAVs using ACO algorithm with Gaussian distribution functions," in *Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on*, 2009, pp. 173-176.

Riadh SAAIDIA received the engineering degree in 2007 from the National Institute of Applied Sciences and Technology (INSAT). He received M.S. degree in 2009 from INSAT, Tunisia and now, he continues D.S. / Ph. D. at the University of Cartage (INSAT), Tunisia in the field of industrial computing and automatic. Between 2008 and 2014, he has worked as a Technologist Assistant in Higher Institute of Technological Studies of kef, Tunisia. His research interest includes multi-robotic systems, artificial intelligence, control theory and computer science.

Mohamed Sahbi Bellamine received the engineering degree in 1991 from the National School of Engineers, Tunis (ENIT). He received his M. S. degree in 1996 and his Ph. D. degree in 1999 both from Ehime University, Japan in the field of Robotics. He is actually occupying the position of Assistant Professor at the National Institute of Applied Sciences and Technology and the position of a visiting Professor at the Georgia Institute of Technology. His research interest includes mainly the Human-Robot Interaction field.

Abdessattar Ben Amor received the engineering degree in 1988 from National Engineering School of Tunis-Tunisia (ENIT) as electromechanical engineer. He received DEA degree, in Systems Analysis and Digital Processing, in 1990, D.S. / Ph. D in 1999 and HDR in 2010, from ENIT Tunisia, in the field of Electrical Engineering. Between 1989 and 2007, he was Researcher at LARA in ENIT. Since 2008, he continues his research at LISI in National Institute of Applied Sciences and Technology (INSAT) in Tunis. He also worked as a Director of university life and Relations with the Environment in INSAT, since 2011. Currently, he is a Lecturer in INSAT. His research interest includes design of electrical machines, automated Systems, linear machines, mechatronics, artificial intelligence, Electronics and robotic systems.

Towards Developing a Cost Effective Solution for Environmental Monitoring

Muhammad Soban Khan, Ans Ali Raza, Zeeshan Musawar, Shoaib Hassan, Taimoor Hassan

Department of Computer Science, COMSATS Institute of Information Technology, Sahiwal[1, 2, 3, 4]

COMSATS Road off GT road, Sahiwal 57000, Pakistan

Department of Computer Science, Lahore Garrison University, Lahore

Abstract

Environment refers to everything that surrounds a person. Environment contains many types of pollution. Most dangerous pollution is air pollution. Most important factor that causes human health is air pollution. Many countries are suffering from air pollution. There are many factors that cause air pollution. Some major factors are smoke, carbon monoxide and high temperature. Many developing countries are creating solutions for detecting and analyzing the air pollution. The main idea of our research is based on proposing a cost effective solution for environmental detection. Our system is a connection between sensors, Raspberry Pi, Microsoft Azure and Android Mobiles. Raspberry Pi gets environmental values with help of Raspberry Pi and sends the data to Microsoft Azure through API, form where Android Mobile gets those values with the help of HTTP request. Our proposed system successfully detected temperature, humidity, hydrogen, methane, propane, carbon monoxide and air level. The results show that our system is most cost effective, secure and easy to use. It will helpful in saving lives.

Keywords: Environment Pollution, Environmental monitoring system, Raspberry Pi, Air pollution

I. INTRODUCTION

Environment refers to everything that surrounds a person. One of the most important part of a person's environment which has great effect on person's health is air. As fresh air keeps one's health good, polluted air can create havocs on a one's health. In last 30 years, scientists and researchers have found a large range of diseases (Asthma, COPD, Lung Cancer etc.) caused by air pollution.

Air pollution has been active from the old times in the form of volcanic eruptions, wildfires and dust storms and due to them, gasses like sulfur dioxide, carbon monoxide etc. are continuously disturbing the atmosphere. In the middle ages, coal combustion/burning as heaters etc. was prohibited in London while Parliament was in session, as it can suffocate people in a closed room. However, the problem of air pollution accelerated because of the increase of emissions of gases and other industrial wastes since the industrial revolution. In modern history, the first recognition that air pollution is more than a local problem arose with a dispute between the states of Tennessee and Georgia in the U.S. in the year 1907. A legal argument started that set the stage for other similar clashes—that were not resolved completely until 1955 federal legislation and the 1970 clean air act amendments. The 1955 legislation only provided for research, but the 1970 act enabled laws for industries emitting poisonous gases and toxic wastes [1].

According to World Health Organization (WHO) 2012 report, around 7 million people died one in eight of aggregate worldwide deaths as an aftereffect of air pollution. The under developing countries in "WHO" which belong to South-East Asia and Western Pacific Regions were found most polluted in the survey results in 2012. In these countries almost 3.3 million deaths happened due to indoor air pollution and 2.6 million deaths happened due to outdoor air pollution [2].

One must know what his/her surrounding contains. Pollution doesn't only exist in outdoor environment but also in indoor environment. Carbon Monoxide is one of the main components of both indoor and outdoor air pollution. Other than that Methane, Propane etc. are also some of the gases which exist in indoor and outdoor pollution. Such gases are affecting the ozone layer. Due to this, the temperature of the world is rising gradually.

Systems should be derived for a normal person to know and understand environmental conditions of his/her

surroundings. So that he/she can understand the dangers of air pollution.

II. RELATED WORK

Now a day's many new products or new versions of old products are being introduced in the market. Developing a new product is a very sensitive matter as customers want to feel easy and comfortable in the usage of the new product as compared to the old product. Many authors think that management of a new product is the combination of art and science. Devising an idea to make a new product is an art whereas converting that idea into a product (according to customer's need and ease) and launching it in the market is science [3]. There are many ideas on air pollution detection which were converted into a product. Some of them are given below.

Computerized tomography technique is one of the previous techniques used to detect air pollution. This technique produces a two dimensional map of polluted area. In this framework, there is a solitary laser source situated at the focal point of the range. This laser beam is rotated and coordinated towards the boundary of the circle. There is a tube shaped reflect so that occurrence laser beam is reflected in a fan beam over point over the circle. The beam from the mirrors is the circular region and strikes a set of detectors lie in same plane parallel to the ground. This method concentrate on lower transmitted laser vitality expanding the reach and capacity to screen the region that contains a few toxin sources [4].

Air Quality Index (AQI) is a scale or index to get the information on the quality of air. It is used by governments to do so. It tells that how much polluted the air is and how much polluted it will become and what health effects will the air have on people by breathing it. It is used by many countries around the world. USA, China, Canada, Malaysia etc. are using such scales. USA Environmental Protection Agency (EPA) calculates the index for five major air pollutants according to the Clean Air Act: ground-level ozone, particle pollution or suspended particulates ($PM_{2.5}$), carbon monoxide, sulfur dioxide, and nitrogen dioxide. Other countries detect the same pollutants as calculated by EPA but also some extra pollutants, such as China also detects suspended particulates (PM_{10}). For all of these pollutants, national air quality standards are established to protect public health. Hourly readings of each pollutant are noted. At the end of each hour pollutants on every site are noted and then measurements are converted into numbers from 0 to whatever the highest range is. The lower the pollutant value is, the better the air quality is and the higher the measurement is, more worse the air quality is. For example, if the AQI scale has range from 0 to 100. Where air quality level lower than 15 means very good, 16 to 30 means good air quality, 31 to 50 means moderate air quality and more than 50 means bad air quality and serious health concerns. [5].

Portable Emissions Measurement Systems (PEMS) measures mobile source emissions, the emissions from combustion engines in cars, trucks, generators and cranes etc. which allow real-world in-use testing. As all these sources have a large impact in air pollution, in 1995 Mr. Breton thought to manufacture a system to detect pollution emission from one of the biggest air pollution production source. It is a modern and innovative system implanted in vehicles to check how much pollution they make. The purposes of PEMS are to integrate advanced gas analyzers, weather station, exhaust mass flow meters, GPS and connection to the vehicle networks. The purpose of PEMS related to air pollution is that it detects pollutants emitted by the engines such as carbon monoxide, carbon dioxide, hydrocarbon etc. together with other engine parameters of the car [6].

III. PROPOSED METHODOLOGY

Unlike the previous environmental systems an easy and cost effective solution is proposed, consisting of:

Hardware

- Raspberry Pi
- DTH11 sensor
- MQ2 sensor
- MQ7 sensor
- MCP 3002

Database

- Microsoft Azure Cloud

User Interface

- Android Application

Raspberry Pi will be core of system and will work as a fully functional computer. Raspberry Pi contains a dedicated processor, memory, and graphics driver for output through HDMI. It does not have an internal memory but an SD card can be used as flash memory. Raspberry Pi needs an operating system to work so, Raspberrian operating system or Windows operating system (Windows IOT) launched in 2015 will be used for Raspberry Pi to work. Figure 1 shows Raspberry Pi 2.

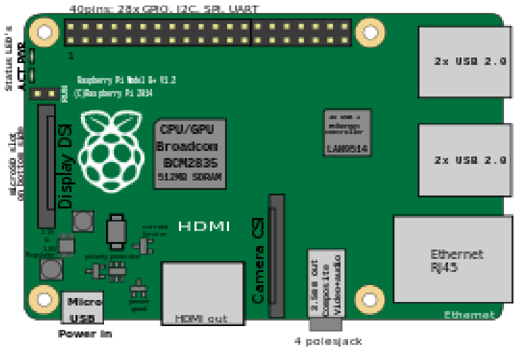


Figure 1

DTH11 sensor will detect temperature and humidity from the surroundings. MQ2 sensor which is a smoke sensor will detect Methane, Propane and Air Level. MQ7 is the most important sensor of the system which will detect the poisonous gas “Carbon Monoxide” (CO). To get the values through sensors, Python, C# or C++ codes will be used. Figure 2 shows DTH11, MQ2 and MQ7 (left to right).

Figure 2

MCP 3002 is an analog to digital convertor. As the sensor retrieve the data in analog form, a convertor will be needed to convert the data to digital form. Figure 3 is a picture of MCP 3002.

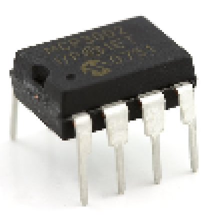


Figure 3

Microsoft Azure Cloud will be used as online database for the system. Data will be sent to Azure Cloud with the help of Restful API because Restful API is not just limited to XML format but can also contain JSON format.

User will need an interface to get his/her desired values. An android application will be used to provide the user his/her desired value from the sensors. The interface must be easy to use as there will be all type of Android phone users to use it.

The benefits of the above components are shown in TABLE 1.

TABLE 1

COMPONENT-BENEFIT TABLE

Component	Benefit
Raspberry Pi	Cheap and small
Microsoft Azure	A secure online database
Android application	Available to everyone
MQ2 sensor	Smoke detection
MQ7 sensor	Carbon Monoxide Detection
DTH11 sensor	Temperature, humidity Detection

All of the components are small, easily available, easily purchasable and easy to use. Thus, making it an easy system to build and provide people environmental reports with ease.

Figure 4 shows the possible architecture of the system.

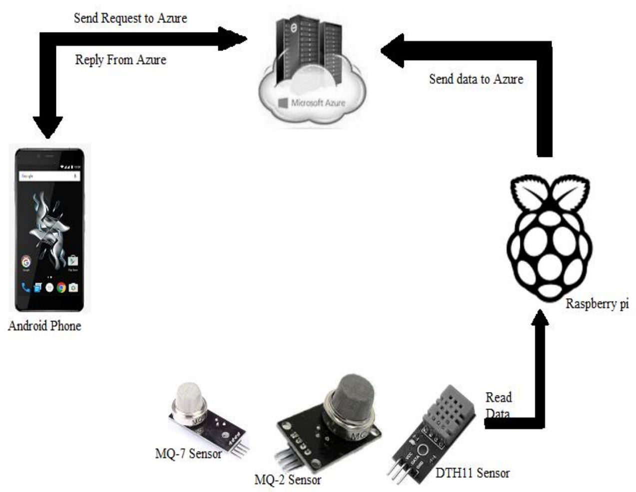


Figure 4

This system is static so can be connected with different modules to perform different environmental actions. For

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

example, the system can be attached to a quad copter to get the data of a wide area.

IV. RESULTS

The system will provide the values form the sensors. The sensors will give the following information collectively:

- Temperature
- Humidity
- Carbon Monoxide
- Methane
- Propane
- Butane
- Smoke
- Hydrogen

Above results can be replaced by replacing the given sensors with other sensors according to requirement.

Figure 5 shows the command in Putty which runs the python code file and shows the results of that code which reads the data from sensors and sends to Azure Cloud.

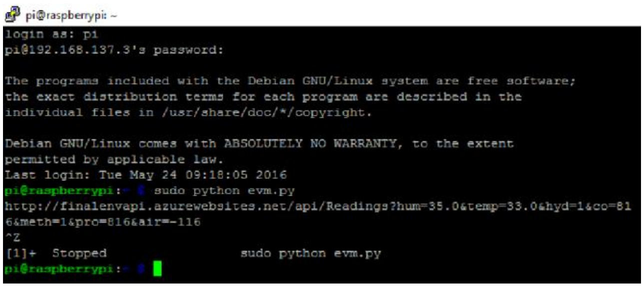


Figure 5

Figure 6 shows the results stored in Microsoft Azure through Restful API in the form of JSON.

Figure 6

Figure 7 shows the front end of the system where user can see the environmental data retrieved from the Azure Cloud with HTTP request.

Figure 7

V. CONCLUSION

Environment refers to everything that surrounds a person. One of the most important part of a person's environment which has great effect on person's health is air. Air pollution has existed for a long time in the form of wild fires and volcanic eruptions etc. Nowadays, it has evolved and become more dangerous due to gasses emitted from vehicles and factories etc. Many systems are active for air pollution detection but there isn't a cost effective system and a system which can provide environmental information to a simple person with the help of an Android phone (which is available to many people now a days). The proposed system is a cost effective solution for environmental monitoring and provides a person an easy way to get environmental information.

References

[1] http://www.idc-online.com/technical_references/pdfs/chemical_engineering/Air_pollution.pdf

[2] http://www.who.int/mediacentre/news/releases/2014/air-pollution/en/

[3] Crawford, C.M. & Di Benedetto, C.A. (2000) New Product Management.

[4] 6th ed. London: Irwin McGraw-Hill

[5] Robert L.Byer, Lawrence A.Shepp, "Two-Dimensional Remote Air-Pollution Monitoring Viatomography".Vol.4/ March 1979 /OPTICS LETTERS.

[6] https://airnow.gov/index.cfm?action=aqibasics.aqi

[7] http://iet.jrc.ec.europa.eu/pems/portable-emissions-measurement-systems-pems

AV Encryption Algorithm to Protect Audio visual Content for IPTV

Muhammad Akram *, C. A. Rahim *, Amjad Hussain Zahid **

*The Institute of Management Sciences (PAK-AIMS), 54660 Lahore, Pakistan

**University of Engineering and Technology, Lahore, Pakistan

Abstract— **Crypt analytical techniques for multimedia technologies particularly audio visual applications have shown some existing flaws while maintaining the security and computational time. This case study is a representative algorithm especially for protection of IPTV contents. The network's reliability and security of contents is the major issue in IPTV media business. The proposed algorithm is the Audio Video MPEG file encryption technique in which the synchronization between audio and video and the frame sequence is shuffled before the transmitting end or vertical device. . The shuffling process is guided by input key frames to point out frame positions. The MPEG video frames are first extracted via spatial pyramid kernel. It divides the stream into regions over different scales and to find out the frame similarity while on merging of AV frames. Then ciphers are implemented to locate the shuffled frames and further genetic algorithm such as AES is used to encrypt. By this way, AV contents of IPTV can be secure from malicious users.**

Keywords— *MPEG, IPTV, CAS, DRM, DES, AES*

I. Introduction

Internet Protocol Television (IPTV) provides digital information and audio video contents using high speed Internet [1]. This IP based managed network is supposed to provide quality of service (QoS) and quality of experience (QoE) with different factors such as security, reliability and interactivity. The Audio Video contents are transmitted or delivered through digital video broadcasting (DVB) by using distributed delivery network, distributed management network and some additional servers [2]. Before transmission, the IPTV contents must be digitized as MPEG format. The requirement of security and copyright is an important issue. The content security, service security and transport security are the major requirements for IPTV setup. The digitized contents can be easily copied. While using computer the audio video clips can be saved and shared rapidly in the same quality as the original. To prevent illegal duplication and distribution of contents, different studies are being conducted. Conditional access system (CAS) and digital right management (DRM) are initial security technologies [3]. CAS has a negative aspect that it cannot provide continuous protection from illegal copy and sharing. Some other approaches for the encryption of AV bit stream are cryptographic algorithms such as DES or AES. The large amount of AV data requires real-time operations particularly in the case of the wireless mobile systems. It is difficult to handle the heavy encryption processing load along with the AV content with respect to computational time. Maintaining computational efficiency based on different performance metrics as encryption decryption time, throughput, CPU process time and memory utilization is very important to provide a continuous protection for a secure system [4].

This study proposes a security algorithm with a technique by shuffling the audio video sync and frames. The shuffling system can be interfaced with distributed delivery network. The content delivery network is based on software and hardware to generate a playlist and play out the contents. The proposed system encrypts the contents for security purpose. This system can capture the contents according to play list and shuffle the audio video contents. The shuffled AV out is fed to content aggregation system. The AV contents are downlinked via satellite receivers [5]. By using such contents which are already distributed by MPEG-2/ MPEG-4 satellites, different signals can be received by integrated receiver decoders (IRDs). The output signal format of IRDs is usually known as serial digital interface (SDI) format.

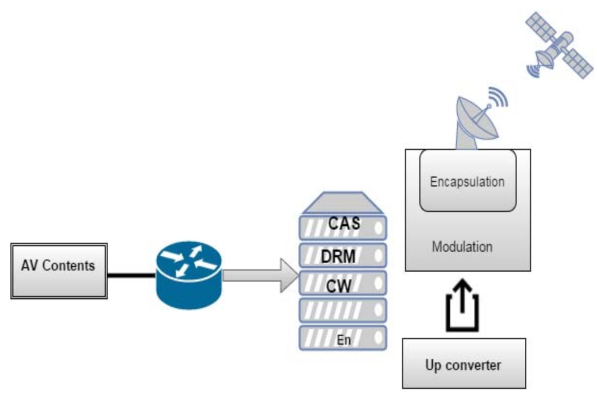


Fig.1a AV content distribution with CAS and Up-linking

The typical AV content distribution with CAS and Up-linking is shown in figure-1-a. In the Figure-1-b the contents are received and router feed these contents to digital subscriber line access multiplexer (DSLAM) which aggregates complex composite signal of AV traffic through multiplexing. It also aggregates the DSL lines over its asynchronous transfer mode or internet protocol network. At the subscriber end there is a reversal system to decrypt or descramble the content. By this

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

way the authentication and the copyright issues can be implemented to IPTV.
This study is based on following; related work will be discussed in section II, proposed system will be described in section III, Implementations and testbed will be described in section IV and Conclusion will be discussed in section V.

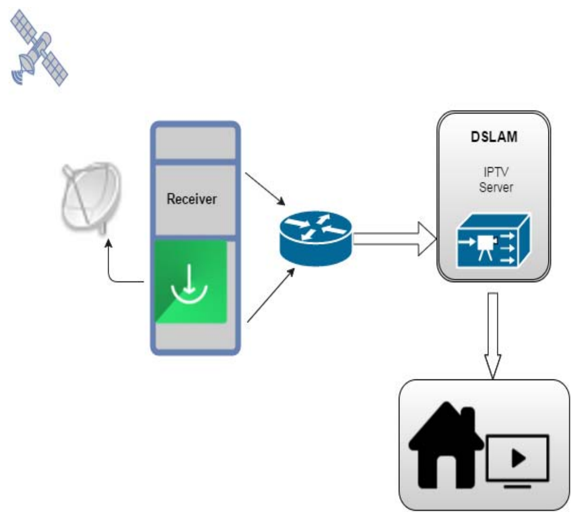


Fig.1-b Satellite-based receiving end for IPTV system

II. RELATED WORK

A. *Protection of digital contents*

The flexible operation and various functionalities can be implemented over the digital contents or data. Digital contents or data can be easily copied and used by authorized and unauthorized users. The identification of original content and copied content is difficult. From security point of view the flexible operations for digital contents can damage the copyright holder's intellectual property. Therefore the technologies for protection of digital contents are needed. The respective security algorithms are such as completely layered encryption, encryption using Permutation [6], selective encryption and Perceptual encryption [7]. The short introduction of these algorithms is given:

1) Fully layered Encryption

In fully layered audio visual content protection schemes, the entire content and bit stream is compressed and encrypted using a standard encryption (DES, AES, IDEA, etc.). In this method let consider ɱ target images (I1 , I2,…., Im). At first only two target images (I1, I2) are stated as first layer input. On the convergence end, two other images ($I1 e^{i\varphi 1}$, $I2\ e^{i\varphi 2}$) are achieved. In the next step two images are considered ($I2\ e^{i\varphi 2}$, $I3\ e^{i\varphi 3}$), this action has to continue till the last image for example $Im\ e^{i\varphi m}$ as the input of the double random phase (DRP) algorithm [8]. This encryption algorithm is not applicable for live and real time applications because of various computation and slow speed [9].

2) Permutation based Encryption

The different permutation algorithms are used to encrypt the content of video as bit permutation, pixel permutation and block permutation. The permutation algorithm [10] has ten steps such as load image, input the 8 bit key, conversion of pixel value into binary, repetition of three planes, rearrangement of the bits with respect to given key, conversion of the permuted to decimal value, transformation of pixels to matrix, Permutation of the pixels with given key, division of the image into 8 blocks and rearrangement of the blocks according to given key.
The scrambling of each and every bit is not necessary. To encrypt video contents, permutation list as secret key is used. After encryption using this technique the video will still be perceptible. With the power of computers, substitution and permutation encryption can be easily cryptanalyzed. The main limitation of permutation based encryption is plaintext attack because the attacker can recover the frames until the scene changes and to update the key frame [11].

3) Selective encryption

In this algorithm the implementation of key generation, encryption and decryption is used [12]. This algorithm has following steps:
Step 1: Key generation: to generate the Key four steps are involved as

i. Selection of prime numbers p¸q
ii. Where n = p * q
iii. lcm (p- 1, q -1)
iv. Selection of integer e

Step 2: Encryption:
(m_χ, m_γ) as Plain text and (e, ɳ) as public key
(c, a) have to send
Step 3: Decryption:

The decryption is based on secret key (p, q, dp, dq) which is received by receiver. So it also takes further steps and the process time [12]. The drawback of selective encryption is that it causes bandwidth expansion and badly impacts on efficiency for compression [13].

B. *Protection technologies for IPTV*

CAS (Conditional Access System) and DRM (Digital Right Management) are the two initial or primary technologies to protect the IPTV contents. The operator feeds an audio video stream to CAS system for allowing only the authorized users to receive certain programs. The major functions of CAS are scrambling and descrambling, entitlement control and entitlement management. By using these features CAS can protect business of paid broadcasting service providers. But this system has some weakness like there is no continuous protection. This system is costly, complicated and hardware based. To manage and control the digital contents from its production, the DRM (Digital Right Management) system is used in the Internet environment. This system distributes the license for terminals [14]. A license consists of a decoding key and usage authority for contents. For issuing and management of the license, the digital contents are encoded by

DRM packagers. A streaming DRM has two types; one is for VOD (video on demand) and other for multi cast contents. This system needs to persevere through failures of system components. Due to such limitations, the DRM health monitoring approvals and keys have to cross check for errors [15].

III. PROPOSED ALGORITHM

A. Initial Discussion

This study proposes AV encryption algorithm with a technique by which both audio and video can be encrypted. The MPEG video clip files as an AV contents require the security mechanism, for this purpose different encryption algorithms were proposed such as fully layered bit stream encryption, pixel block permutation encryption and selective key generation encryption algorithm. However there are some limitations and drawbacks while implementing these algorithms particularly for real-time streaming. The AV content delivery for IPTV services requires an intelligent network with advanced security to protect the confidentiality, integrity and availability of videos. The distribution of AV contents over IP can be subject to the hackers, threats and vulnerabilities which may lead to end users dissatisfaction. For this reason, AV encryption algorithm and its implementation setup is proposed in this study. The proposed algorithm is the Audio Video MPEG file encryption technique in which the synchronization between audio and video and the frame sequence is shuffled or scrambled before the transmitting end or vertical device. The MPEG video clip file has a fixed size header with different information such as

I. Clip file number
II. Image width
III. Image height
IV. Frames rate per second

The MPEG video clip file format is shown in figure 2.

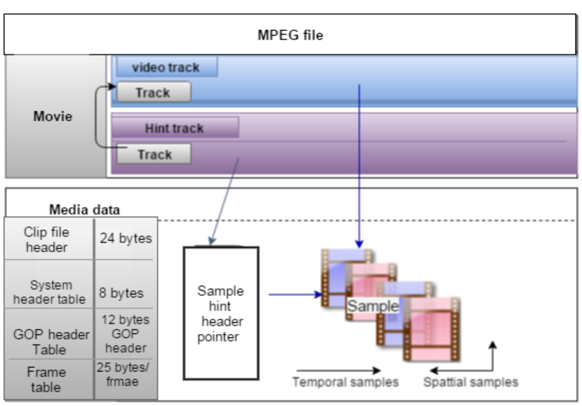


Figure.2 MPEG video clip file format

In the MPEG data, the record of individually system header is composed by system header and each record at starting point has byte offset. The GOP header table is based on index of the system header and this index is very basic character for GOP. The entrance of each frame in the file is composed in the frame table with header indexing; this indexing has following information such as

i. The table of the frame's header
ii. The number of byte from the starting position of the clipfile frame by frame)
iii. The frame size (I, P, or B)

As clip file header, 24 bytes are added as the file representation while eight bytes for each system header and GOP header has 12 bytes. So for the overall frame 25 bytes are used [16]. MPEG data structure or layers has three types of picture or frames which are also known as GOP, these frames are

Intra-coded (I) frame
Predictive-coded (P) frame
Bi-directionally predictive-coded (B) frame.

These frames are also called reference frames and two of them are dependent frames which are P and B frames, and I frame has no reference to other frames. The P frame is coded by motion compensated estimation from a previous reference frame and the B frame is coded from a past or future reference frame using motion compensated prediction. The MPEG video structure with respect to GOP is shown in figure 3. Each GOP has a specific frame order such as display order and coded (transmission) order, in display order the input is given to the encoder and output from the decoder. In coded transmission order the output is taken from the encoder and the input is fed to the decoder. The GOP frame order meaningfully determinates the video and coded quality on the same bitrate. The available bitrate is reduced if only I frames are used. The video quality can be improved on the increment of P and B frames.

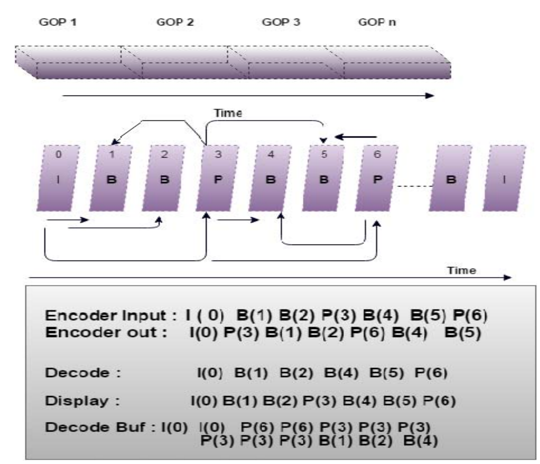


Figure.3 MPEG- video architecture

The video sequence header of GOP supports for three basic functions such as random access, fast search, and editing. The compressed frame ordering for decoding with respect to the presentation timestamp (PTS) and decoding time stamps (DTS) is shown in figure 4.

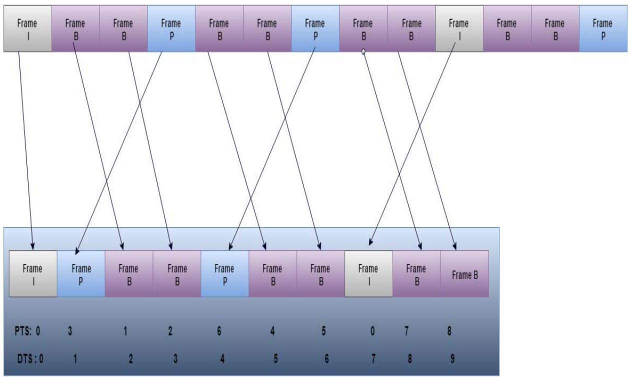


Figure.4 MPEG frames order for decoding [17]

In order to restructure the bidirectional B-frame from the previous I and P, both must arrive first and must be different to the order for appearance on the TV screen. The Decode Time Stamp indicates a specific time to decode the frame in a specific order. The Presentation Time Stamp indicates a specific time to display the frame with the help of an embedded clock for time reference [17]. These frames are used to propose the AV encryption algorithm in which the AV frames along with time sequence are shuffled with shuffling algorithm.

B. *Shuffling Algorithm*

Permutation based re-ordering from primary variation into the randomly arranged data elements of a sequence and to scramble data elements into a random placement is called shuffling. A shuffle algorithm is the logical converse of a sorting algorithm, for example a permutation based sequence data D of A archives A0, A1, ..., An-2. The sequence data D is measured as shuffled for a possible selection k of any archives Ai and Aj where i $\neq$ j, that the probability of Ai $\leqslant$ Aj is equal to Ai > Aj for 0 < k $\leqslant$ N. For the kth possible selection, the probability þk is:

þk(Ai . Aj) = þk(Ai > Aj) where i $\neq$j and 0 < k . N þk(Ai . Aj) = þk(Ai > Aj) where i $\neq$j and 0 < k . N:

1. Divide an array in to 5 parts and each part have same number elements
2. The position of elements of each part is shuffled in descending order
3. Check if the position of some element is still same after the shuffling then reshuffle its position with right neighbour
4. Check if reshuffled elements numbers are in sequence then shuffle its position with grater element
5. Compare the resultant array with the input array and mark the shuffled positions

The above mentioned shuffling can be explained by the figure 5 which is an example of shuffling the position of 25 elements of an array.

The shuffling positions of resultant array according to the figure 5 can be explained by the table 1.

According to the Table.1, each element of input array is shuffled with the help of shuffling algorithm in which the array is divided into 5 equal parts and then replacement of the positions of each element in a random descending order which is the requirement of this algorithm is conducted, the replacement positions of each element is shown in figure 6.

Now the resultant random descending order array is marked with the replacements so that it can be rearranged in the ascending order as the input array. To convert the descending order array in to ascending order the merge sort algorithm is used which is the reverse of shuffling algorithm.

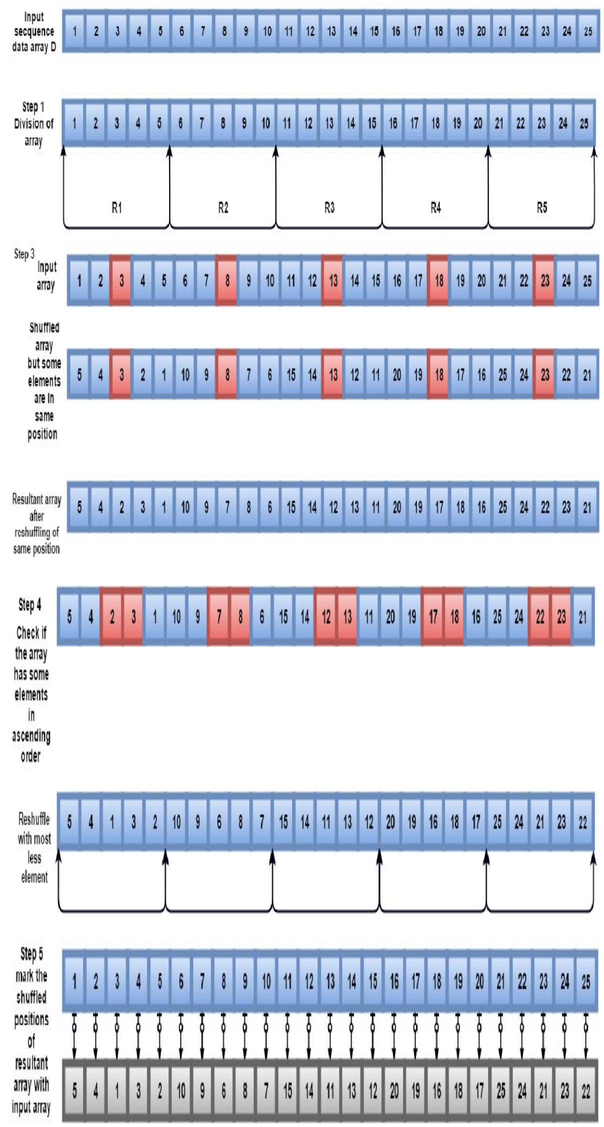


Figure.5 Implementation of shuffling algorithm

In merge sort algorithm the array is divided in to 5 parts and then recursively sort each part. The recursive structure for merge sort is based on three basic functions first one is divide

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 6, June 2016

second one is conquer and third one is combine. In the divide function, divisions of the task into subtasks are similar to original in the size. In the conquer function, conquer the substitute tasks recursively. However if these tasks or problems are still solvable, then these can be solved in a straight forward manner. The final function is combined. In combine function, the solutions are joined to create a solution for the original task or problem. To achieve the original array (input) the following steps are given according to the merge sort algorithm as given:

1. Divide the input array of 25 element into five subsequence of 25/5
2. By using merge Sort, recursively sort the resultant five arrays which are subsequence
3. To produce the array in ascending order, merge the five sorted subsequence

Table 1: Resultant shuffled array

Input Array	Resultant Shuffled Array
1	5
2	4
3	1
4	3
5	2
6	10
7	9
8	6
9	8
10	7
11	15
12	14
13	11
14	13
15	12
16	20
17	19
18	16
19	18
20	17
21	25
22	24
23	21
24	23
25	22

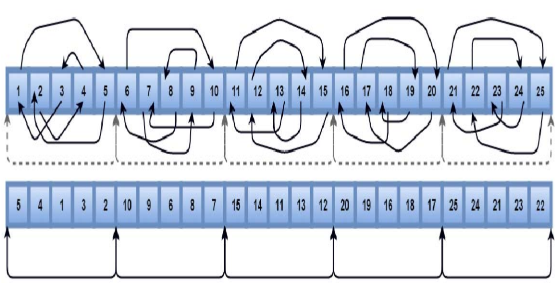


Figure.6 Shuffling Positions

The first step is to divide, second step is to conquer and third step is to combine so by this way the shuffled array can be rearranged in ascending order or the input array. The steps of merge sort are based on to rearrange the shuffled array and shown in the figure 7.

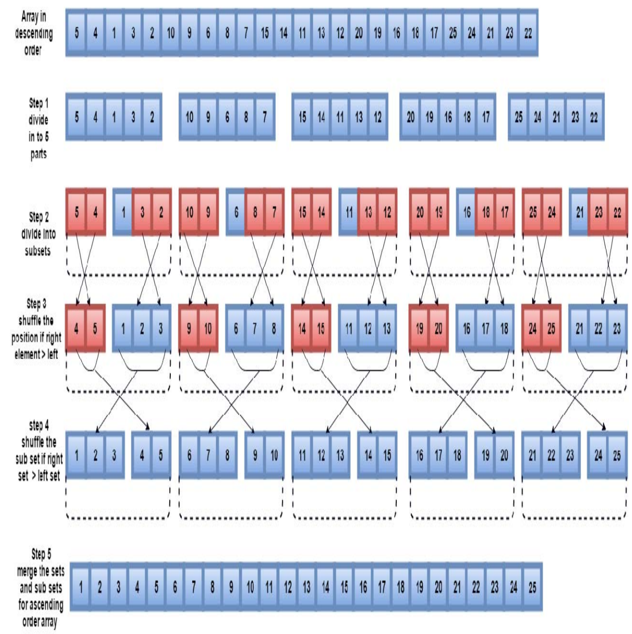


Figure.7 Reshuffling for ascending order array by Merge sort

C. Implementation of shuffling algorithm for AV Encryption

The encryption is the transformation of original information (in any form such as data) into a secret code for an effective data security. On the other hand the end user can read or receive an encrypted information or data with a secret key or password. The shared secret key or password allows the authorized end users to decrypt the encrypted information. For this, shuffling algorithm is used to propose the AV encryption algorithm. The AV contents are distributed via TV channels through transmission setup, in which the playout software plays the content list and encodes these contents to feed the earth station for further distribution via satellite. So in this scenario the AV contents can be encrypted before playout, and then encrypted AV contents are ready to transmit via satellite or any other distribution setup which may be wired or wireless. The difference between the general setup and the proposed AV encrypted setup is shown in the figure 8. The AV contents in the MPEG format are played in a unique standard, as the video displays a set of pictures in which the frames are formed speedily at a particular rate while recording, capturing or transmitting the AV content.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

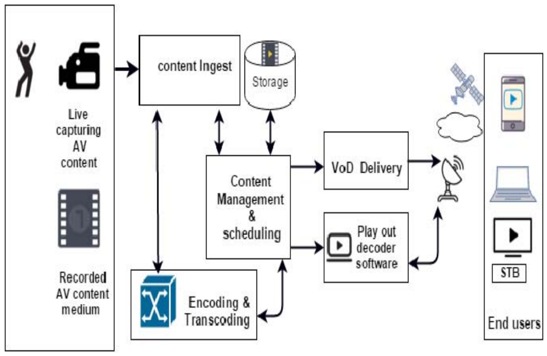


Figure.8 Formal TV transmission setup

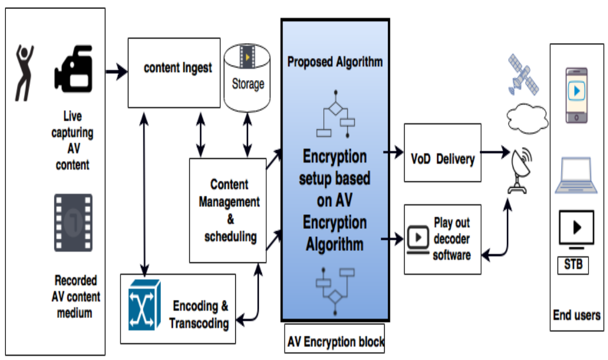


Figure.9 proposed encryption setup

The end user can watch these transmitted or distributed AV content on television screens or over mobile screens via satellite, cable network, web or any other distribution setup. The characteristics of video stream are

i. Speed of frames with respect to frame the number per unit time
ii. Video resolution (size of video according to Pixel)
iii. Interlacing (scanning method, consecutive numbers to each horizontal scan lines divided in to fields)
iv. Colour space bits per pixel (bpp)
v. Bit rate to represent a video or audio

D. Importance of Frame rate for AV Encryption

For AV encryption the main focus of this research is on:

I. Frame rate
II. The quantity of frames per unit time
III. Number of frames per second- fps.
IV. Frame sequence to display a picture
V. Audio Video frame synchronization
VI. Frame positions

There are different standards which are based on above mentioned characteristics of AV frames such as:

a. Phase alternation line (PAL)
b. Sequential colour with memory (SECAM)
c. National Television System(s) Committee (NTSC).

PAL and SECAM standard has working principle on 25 fps and NTSC working principle on 30 fps. Every frame is an orthogonal bitmap to make a raster of pixels, which is actually the frame size with frame width (W) and height (H) as frame size = W*H, normally 680*480 pixels per frame is the size of frame. One pixel is an element of a frame which is a property such as hue colour, contrast or brightness which is represented by a fixed number of bits, in case of 25 fps the bits per frame can be calculated as (W*H)*25.
The frame rate can be varied with respect to the capturing device like camera and the motion of the object which is captured for video. 25f is a progressive format and runs 25 progressive frames per second to achieve cine motion artefacts. Frame rate is variable with respect to the object motion, capturing device and the output display screen and to reduce the shutter rotation. Shutter rotation is the blank portion between two frames while playing the audio visual content. The moving pictures particularly MPEG is based on frame rate and the persistence of vision by which a human eye can experience moving images made up of individual frames on a film strip. The apparent motion represented in any medium such as films, TV, multimedia computer, etc. is displayed by a series of sequentially still images in a speedy sequence. As shown in figure 10 the frame rate per second.

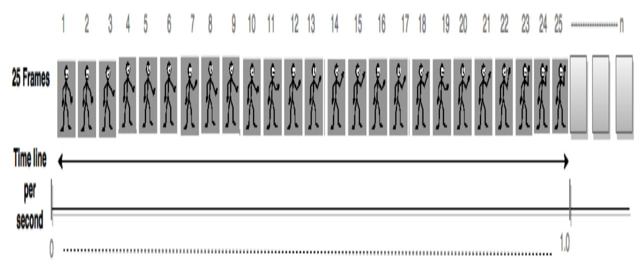


Figure.10 Frame rate number of frames per second

E. What happened if the positions of frames per second are randomly shuffled?

While playing out AV contents, the frame rate per second is mentioned in figure 10 in which the time line shows the number of frames and their position in one second for streaming. According to the figure 8 the AV contents are ready to play out for transmission after content management and scheduling but without any security mechanism. It is important to mention here that what happened if the positions of frames per second are shuffled to encrypt the AV stream before transmission. To shuffle the frames positions of each second, it is required to cut each frame from a stream and then shuffle their position randomly. This setup is shown in figure 11, in which the audio video cutter is used to cut and shuffle the frame positions of AV content stream before transmission or distribution.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

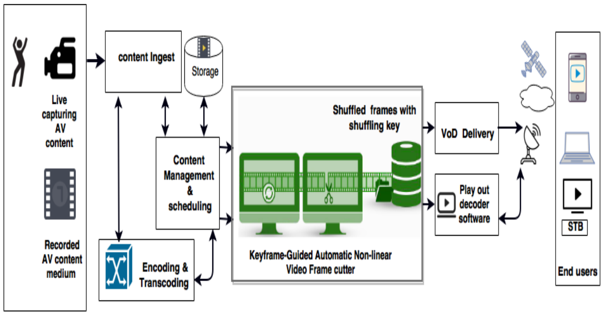


Figure.11 Proposed setup for frame cutting and shuffling

To cut and shuffle the AV contents with the help of proposed algorithm the following steps are used

i. Capture AV contents from content management and scheduler
ii. Extract the frames or division of video in to frames
iii. Check the number of frames per second
iv. Cut the frames and mark their actual position
v. Shuffle the positions of each frame per second
vi. Record the shuffled frame positions as key

By using the above mentioned steps the AV encryption can be achieved. To understand how the frame positions can be shuffled, let assume that one second of AV stream has 25 frames which are divided in to 5 equal parts by video cutter tool. The positions of the 5 frames with respect to the input stream can be presented in the following matrix.

	Number of frames				
Frame position	f1 (a)	f2 (a)	f3 (a)	f4 (a)	f5 (a)
	f1 (b)	f2 (b)	f3 (b)	f4 (b)	f5 (b)
	f1 (c)	f2 (c)	f3 (c)	f4 (c)	f5 (c)
	f1 (d)	f2 (d)	f3 (d)	f4 (d)	f5 (d)
	f1 (e)	f2 (e)	f3 (e)	f4 (e)	f5 (e)

Figure.12 Input Frame numbers and their positions

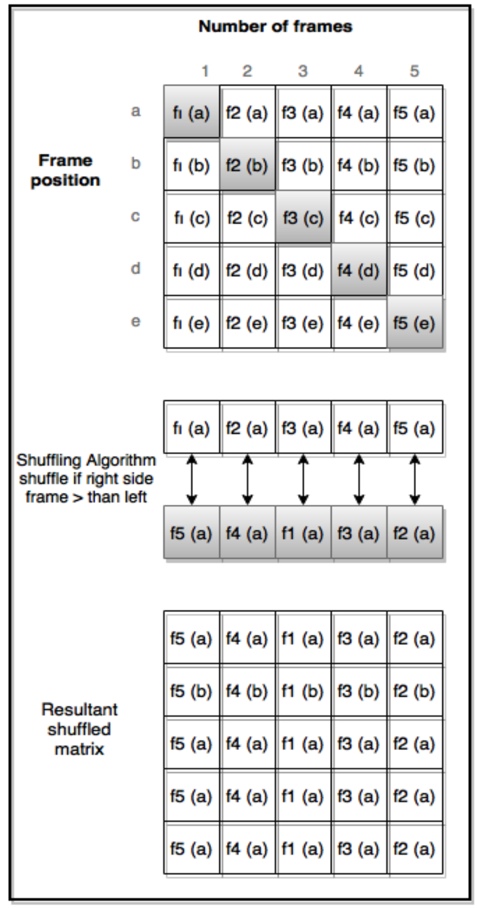


Figure.13 Frame shuffling block

Figure 12 shows the division of video stream in case of 25 frames per second, according to algorithm this video stream can be divided in to 5 equal parts containing 5 frames in a sequence. To record the frames number and their positions, a matrix is shown, in which a, b, c, d, and e are the frame positions and f1, f2, f3, f4 and f5 are the number of frames.

After the division each part of the video stream is fed to the shuffling block which is shown in figure 13. The shuffling of each frame position is conducted by the shuffling algorithm where the position of one frame is shuffled with other by using the shuffling algorithm condition such as, if the right side frame is greater than left side frame, then shuffle its position.

F. AV Encryption Algorithm flow chart

Each frame of a video stream can be shuffled to achieve the resultant shuffled matrix along with the record of actual position of frames to generate a key. In the case of 1st five frames of 25 frames per second, the key is resultant shuffled positions with the mark of actual positions of frames. The AV encryption algorithm flowchart is shown in figure 14.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

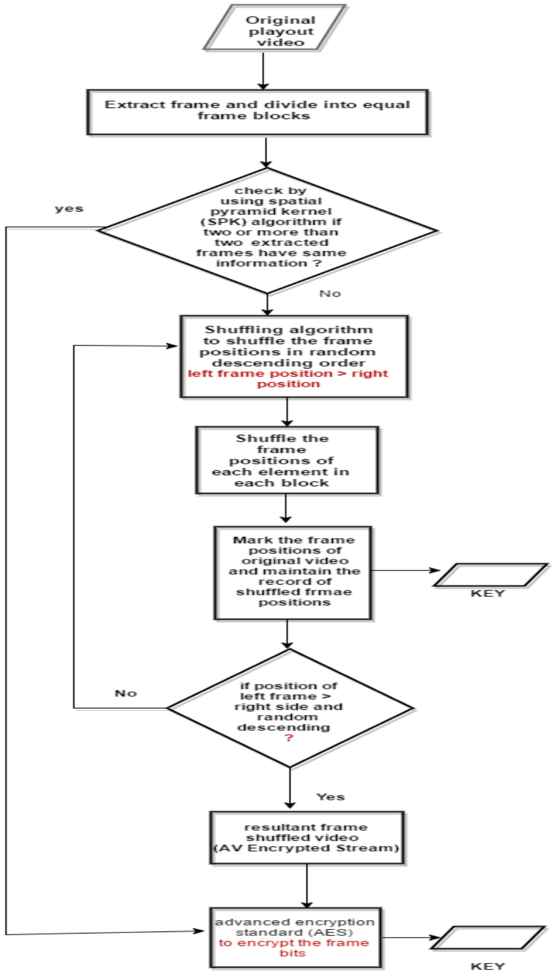


Figure.14 AV Encryption Algorithm flow chart

According to algorithm flow chart the first step is to extract the frames and divide into equal frame blocks. Before shuffling the positions of each frame it is important to check if two of more than two frames have same information.

The similarity of the frames from each divided block can be checked by spatial pyramid kernel (SPK) algorithm. This algorithm is used to divide the AV stream frames into random blocks or regions over different scales, extract interest point descriptors (dense scan resolutions or hue scan), build spatial histograms and create intersection kernels. And if yes some frames are similar with each other, then these frames are directly fed to AES encryption stage to encrypt bit levels of frames. And if the frames which have not the same information but different information then the positions of these frames are fed to shuffling stage to shuffle their positions in random descending order such as the left frame position > right frame position. This step shuffle each frame position in each block and then mark the shuffling positions with the actual positions of the frames to generate the key. While maintaining the key, it is also checked if some frame's positions are still in ascending order, then repeat the third iteration of algorithm to apply reshuffle the frames. And if each frame position of a whole stream is randomly shuffled in descending order then the resultant of shuffling algorithm AV encrypted stream is fed for next step to provide further security via advanced encryption standard (AES). To encrypt the original AV play out stream the position of each frame is shuffled by using shuffling algorithm, where the Key is the number of shuffled positions which is used to decrypt the AV contents for end-users.

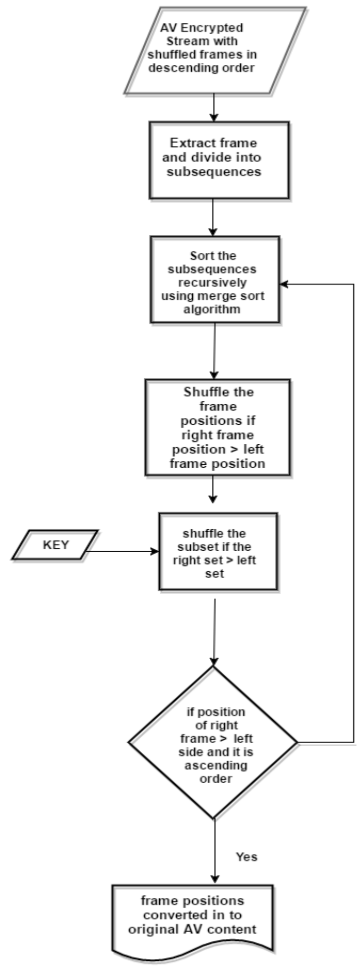


Figure.15 AV Decryption algorithm flow chart

The decryption of frame shuffled AV contents can be possible by using sorting algorithms; in this case the merge sort is used while using the shuffling frame key. The decryption flow chart is shown in figure 15 in which the Key is actually the shuffled position of frames to match with original position. At the receiving end or the end user's setup the encrypted AV contents in form of shuffled frames in descending order can be decrypted by merge sort algorithm with the help of key.

The shuffled in descending order frames of AV contents is considered as a descending order array which have to sort for ascending order as an original array, for this the array is divided into subsequence then shuffle the frame positions with the help of shuffling key. Basically the merge sort algorithm is

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

used to decrypt the randomly descending order frames of AV contents.

G. Block diagram of proposed AV encryption setup

The block diagram of overall proposed algorithm for AV encryption and decryption is shown in the figure 16.

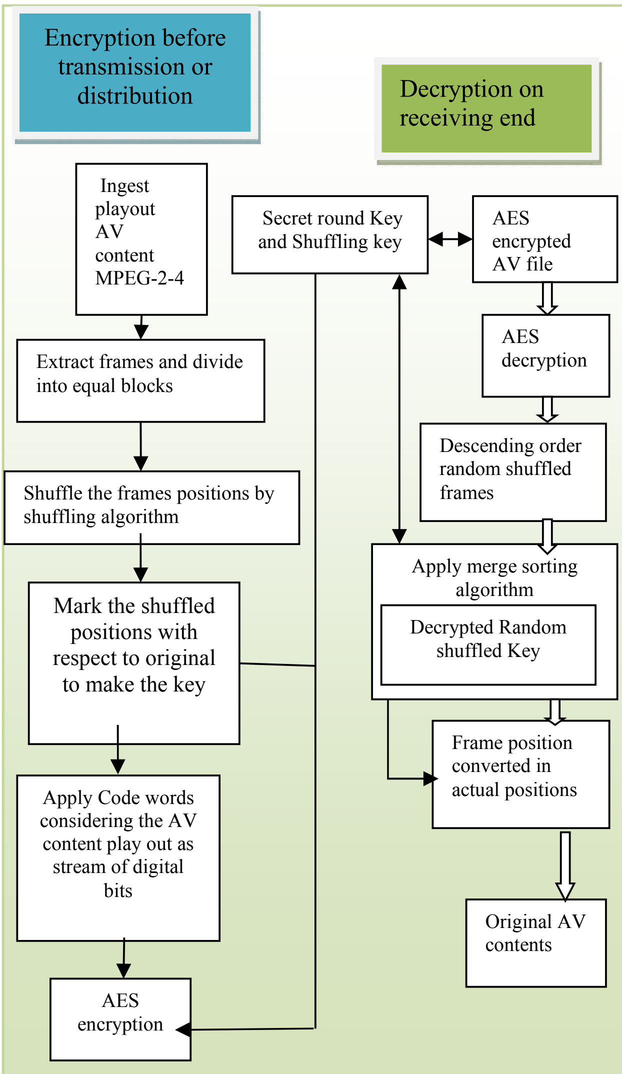


Figure.16 Block diagram of proposed AV encryption setup

To encryption the AV contents before transmission or distribution, the frame positions are shuffled with the help of shuffling algorithm which can be practically implemented to secure the AV contents, and for further security while transmitting or distribution of these contents the formal encryption algorithms such as advanced encryption standard (AES) which is a symmetric-key block cipher. AES encryption uses four types of transformations such as substitution, permutation, mixing, and key-adding for the encryption of 128 bits with its different iterations called rounds. AES encrypts the bits of each shuffled frames further for more security. At the receiving end the AES encryption can be decrypted with the help of random key. The AV encryption algorithm at the receiving end can be decrypted with the help of shuffling key. So by this way the AV contents in the form of MPEG format can be double secured by AV encryption algorithm and the AES implementation to protect the confidentiality, integrity and availability of videos. The analysis of proposed AV encryption and its decryption algorithm is given as that in case of AV encryption with implementation of shuffling algorithm.

H. Analysis of Proposed AV Encryption Algorithm

Shuffling is an unsort algorithm as the logical inverse of a sorting algorithm however it has a probabilistic relation among all the data elements.

Let an array with 25 elements (same as 25 frame per second the AV frame rate) in a sequence with ascending order and each element has to shuffle for the specific position such according to following steps:

i. Divide an array in to 5 parts and each part have same number elements
ii. The position of elements of each part is shuffled in descending order
iii. Check if the position of some element is still same after the shuffling then reshuffle its position with right neighbour
iv. Check if reshuffled elements numbers are in sequence then shuffle its position with grater element
v. Compare the resultant array with the input array and mark the shuffled positions

The initial call is:

shuffle(0, data.length-1, 0, data, indx);

The call to the method shuffle, with the sub-array from 0 to 24 elements (frames) length of the data array, an initial index position of 0, the data sub-array (the initial array passed), the index of frame positions, the frame schedule, and the number of frames to use in shuffling the array. To illustrate the initial first step to partition the array into sub-arrays is used for illustration. So the array is divided in to 5 equal parts containing equal elements. The all the shuffling steps one by one is implements; overall the shuffle is in performance linearly proportional in both time and space, or O(n). The space performance, or memory utilized by the shuffle algorithm is the simplest analysis. The shuffle uses the data of an indices and array elements (frames) schedule, along with the number of frames to use, and the starting index position. The key, frame schedule, and number of frames to use are read-only constants never updated. The index position is updated, as are the sub-array boundaries. The starting array of data elements and following sub-arrays are the same array structure, passed recursively. Hence the size of data elements is N, and the frame schedule and key are of size N. The other parameters passed, and used within the shuffle are a constant number p. The performance complexity is basically the sum of the arrays and the variables. The three arrays (input array, divided array or subarray and shuffled array) are of size N,

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

and thus in Big-O notation is p×N. The space complexity is the arithmetic expression p×N+p. Using Big-O notation; the space analysis performance is O(p×N+p) that is linear complexity. The time performance or runtime performance of the shuffling algorithm is more involved because the partitioning recursively into sub-arrays recursively has the possibility of complex mathematical analysis. Before delving deeper into the analysis of the time performance of the shuffle, the three points that relate to time performance are:

I. Each frame or element at an index is used only once to shuffle.
II. Each data element for a frame is accessed only once.
III. For a given data element frame position of, only s frames are used where left element or frame position of an array > the right side frame

For an array of N-elements, there are a constant c number of frames to use, the given position s of frames. Each frame extracted is used once in the shuffle, and a data element has its frame accessed once. Thus the time performance is a product expression of the number of frame accesses a, the number of frames used p, and the number of total elements N. The time complexity is expressed in the form of an arithmetic expression of:

$$T = a.p.N$$

Since each frame is accessed once for each pass to shuffle the elements, the time performance expression is simplified to:

$$T = 1\times p\times N = p\times N.$$

Thus the time performance is linear or O(p×N) or more simply O(N). Analysis of the time performance is that the array of N-elements and using s-Frames to shuffle in descending order randomly. The size of N after shuffling remains same, but the positions of frames s are shuffled p. The algorithm accesses the N ´ p positions of the new resultant array. Thus to access all the frames in the array once will have a time performance complexity of O(N) which is also equivalent to O(N×p) or on other hand O(p×N).

On the decryption side the merge sort algorithm is used to rearrange the shuffled frames position to generate the original array. In merge sort algorithm the array is divided in to 5 parts and then recursively sort each part. To achieve the original array (input) the following steps are given according to the merge sort algorithm as given

I. Division of 25 sequence elements which are sorted into five subsequences of 25/5
II. Recursively sort the five subsequences via merge sort
III. To produce the array in ascending order , merge the five sorted subsequences

In the light of above mentioned steps, there are 2 inputs as sequences with whole length. Let the length be n elements and it should be moved its each element to the output. The merge time is O(n) for this algorithm. The number of elements is N and the number of levels is presented with big O such as O(logN) for each level it is O(N), so the over-all level for sorting is O(N*logN).

IV. Test bed tools Results discussion

A. Encryption algorithms results comparison

The comparison results between fully layered, permutation, selective and AV encryption algorithm is verified according to various performance parameters such as

I. Encryption ratio (ER)
II. Speed (S)
III. Compression (C)
IV. Standard Format (SF)
V. Encrypted security(ES)

The given table 2 shows the comparison with different levels of observation such as high (↑), low (↓), variable (var), average (Avg), satisfied (√) and not satisfied (°√).

Table 2: Encryption algorithms results comparison

Encryption Algorithm	ER	S	C	SF	ES
Fully Layered	100%	↑	√	Yes	°√
Permutation	100%	↑	°√	Yes	°√
Selective	var	Avg	°√	Yes	°√
AV	100%	↑	√	Yes	√

B. Tool for Frame extraction and shuffling time

To implement the proposed algorithm in which the frames have to extract from playout stream, the virtual dub as an open tool is used. To calculate the speed of frame extraction from a playout stream, a 27.240 second duration MPEG-2 video clip is tested for the validation of proposed algorithm. The snapshot or print screen of the 27.240 second MPEG-2 audio video clip is shown in figure 17. The 27 second duration video has frames with the ratio of 25fps which are 25*27.240 = 681 frames.

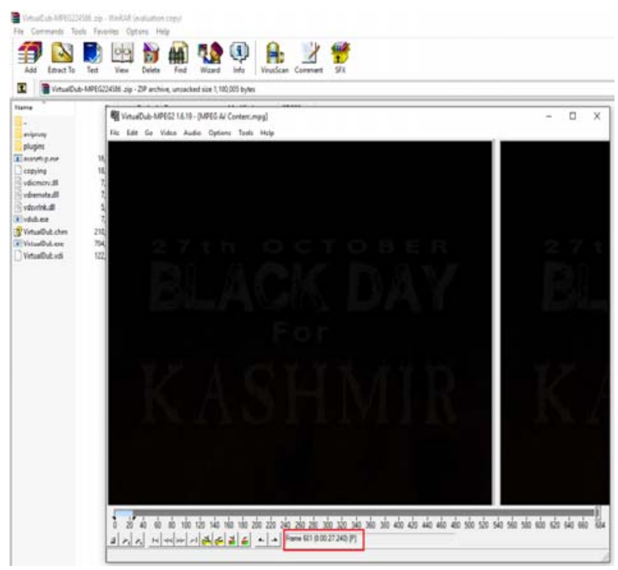


Figure 17. Calculation of number of frames and frame's position

Figure 18. Frame extraction and shuffling time

To implement the proposed algorithm step 2 which is frame position marking and then extract frames from given AV stream, virtual dub is very useful as a testbed. The extraction of frames to shuffle their actual positions, the virtual dub has an option to extract the frames, to conduct a test only 25 frames of one second video are extracted which takes only 0.01 seconds as shown in figure 18 of testbed snapshot or print screen. The extracted frames are then saved in shuffling block to generate a new video with shuffled frame positions called encrypted video; however the actual frame positions are marked for decryption at end user.

V. CONCOLUSION

The IPTV will rapidly be available to the customers as it is being commercialized and vitalized. However, the security issues to protect the digital contents and provide authentication is complicated. To address these complications there are many systems has been designed. The CAS and DRM are to security technologies but these have some weaknesses as a system is too complicated and costly. The proposed algorithm is a simple technique which is not much complicated and it is designed to address both the protection and authentication of AV contents. The functions of proposed algorithm are as randomized that it is not possible for hackers to know that which functions where used to encrypt the AV contents, and also the key length is much higher making the technique an efficient one. As for the future work a guided video cutter or editors for live streaming software is required in regard to the authentication method for the devices with poor calculation ability. Working on MPEG frames extraction via spatial pyramid kernel to split AV stream's frames into regions over different scales and to find out the frame similarity helps the TV professionals to design a guided automatic AV content editing tool. To compute the similarity between extracted frames the dense SIFT features may be used. The future work to extend the proposed algorithm is to almost all types of AV streams rather than IPTV to secure and increase the efficiency by reducing the running time.

REFERENCES

[1] Broadband Services Forum,IPTV Explained-part 1 in a BSF Series,Fremont,CA, www.broadbandserivcesforum.org.

[2] Digital Video Broadcasting (DVB); Transport of MPEG-2 TS Based DVB Services over IP Based Networks, ETSI TS 102 034 V1.5.1 (2014-05) [Online]. Available: http://www.etsi.org/deliver/etsi_ts/102000_102099/102034/01.05.01_60/ts_102034v010501p.pdf

[3] Digital TV Group, conditional Access (CA) Tutorial, http://www.dtg.org.uk/reference/tutorial_ca.html.

[4] http://www.ijcttjournal.org/Volume1/Issue-3/IJCTT-V1I3P107.pdf

[5] ISO/IEC DIS 13818-2,Infromation Technology-Generic Coding of Moving pictures and Associated Audio Information : Video international organization Standardization (ISO) .

[6] Kalpana Singh and Shefalika Ghosh Samaddar, "Enhancing Koyama Scheme Using Selective Encryption Technique in RSA-based Singular Cubic Curve with AVK", International Journal of Network Security, Vol.14, No.3, PP. 164–172, May 2012 [Online]. Available: http://ijns.jalaxy.com.tw/contents/ijns-v14-n3/ijns-2012-v14-n3-p164-172.pdf

[7] http://www.iosrjournals.org/iosr-jce/papers/Vol16-issue3/Version-4/A016340105.pdf

[8] Ayman Alfalou, Ali Mansour "A new double random phase encryption scheme to multiplex and simultaneous encode multiple images," Applied optics, Optical Society of America, 2009, 48 (31), pp.5933-5947. <hal-00579204> [Online]. Available: https://hal.archives-ouvertes.fr/hal-00579204/document

[9] Wireless Networks and Computational Intelligence Communications in Computer and Information Science Volume 292, 2012, pp 599-604

[10] Avi Dixit, Pratik Dhruve and Dahale Bhagwan, "Image Encryption using Permutation and Rotational XOR Technique," Computer Science & Information Technology (CS & IT), vol.06, pp. 01–09, 2012.

[11] D.Socek, S. Magliveras, D.Culibrk, O. Marques, H.Kalva, B.Furht, "Digital Video Encryption Algorithms Based on Correlation-Preserving Permutations", Journal EURASIP Journal on Information Security, Article No. 10, 2007.

[12] Kalpana Singh and Shefalika Ghosh Samaddar, "Enhancing Koyama Scheme Using Selective Encryption Technique in RSA-based Singular Cubic Curve with AVK", International Journal of Network Security, Vol.14, No.3, PP. 164–172, May 2012.

[13] Soumya Balachandran, Sangeeta Sharma," Selective Encryption of Plaintext Using Multiple Indexing", IOSR Journal of Computer Engineering, Volume 16, Issue 3, Ver. IV ,2014, PP 01-05.

[14] Shigue Lian. "DRM for the home TV based on Scalable Video coding", IEEE Trans.on Consumer Electronics,Vol.54,2008

[15] K.Kerpez, D.Waring, G.Lapiotis, J. Bryan Lyles, Ravi Vaidyanathan, "IPTV Service Assurance", IEEE Communications Magazine, 2006.

[16] Lawrence A. Rowe, Ketan D. Patel, Brian C Smith, and Kim Liu, "MPEG Video in Software: Representation, Transmission, and Playback", High Speed Networking and Multimedia Computing, IS&T/SPIE Symp. on Elec. Imaging Sci. & Tech., San Jose, CA, 1994

[17] D.Isovic, G.F.Liesbeth Steffens, "Timing constraints of MPEG-2 decoding for high quality video: misconceptions and realistic assumptions", IEEE 15th Euro micro Conference on Real-Time Systems, 2003

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Secure Speaker Biometric System using GFCC with Additive White Gaussian Noise and Wavelet Filter

Gaganpreet Kaur
Research Scholar
Deptt. of CSE,
I.K. Gujral Punjab Technical University, Punjab, India

Dr. Shashi Bhushan
Professor
Deptt. of Information Technology,
Chandigarh Engineering College,
Landran, Punjab, India

Dr. Dheerendra Singh
Associate Professor
Deptt. of CSE,
CCET, Sector-26,
Chandigarh, India

***Abstract*—Speaker Identification (SI) aims to identify the speaker's identity from the given list of speakers. Speaker identification is efficient under the clean training and testing environment conditions. In real environment application, there occurs mismatch between training and testing environments due to background noise, which degrades the system's performance and security. So, robust speaker identification is the important issue in research. This paper describes the recently used front end algorithm based on Gammatone Frequency Cepstral Coefficients (GFCC) along with speech detection algorithm and Cepstral mean normalization (CMN). System makes model using Gaussian Mixture Model (GMM) Classifier, which uses iterative Expectation Maximization (EM) Algorithm to estimate the Gaussian model parameters. Training data is taken in clean environment and all test utterances are corrupted by adding White Gaussian Noise (AWGN). This paper aims to improve the robustness of speaker identification even when additive noise is added during testing phase. For improvement Wavelet Filter is implemented to de-noise the speech signal. Experiment is carried out in real database oriented and stored database oriented relative to the Attendance System application. Experiment is carried on 100 speakers saying phrases like 'Yes mam' "present mam", 'Yes sir', 'present sir' with 4 types of utterances for each phrase (so database includes 400 utterances). Experiment results obtained shows better performance in noisy environment. The results for stored database oriented experiment show that the algorithm gives 85% of Correct Recognition Rate (CORR) while using wavelet filter and 73% without using the filter. The results for real database oriented experiment shows 74% of identification rate while using wavelet filter and 45% without using the filter.**

Keywords— Gammatone Frequency Cepstral Coefficients (GFCC); Gaussian Mixture Model (GMM); Cepstral mean normalization (CMN); Robust Speaker Identification, Additive White Gaussian Noise (AWGN); Wavelet Filter.

I. INTRODUCTION

Biometrics refers to the process of identification of humans with the help of their traits which can be physiological or behavioral [1]. The physiological characteristics are related to shape of the body like Fingerprints, Palm Veins, Face recognition, DNA, Hand Geometry, Iris Recognition, Retina, etc [2]. The behavioral characteristics are related to individual's behavior like gait, voice/speech, typing rhythm, etc. The selection of biometric trait for any application depends upon the characteristics of trait and user requirements. There are numerous Biometric applications. This paper discusses about an application of Speaker Identification in Attendance System, which includes Speech Biometric Trait.

Speech is a simple and first medium among human beings to communicate with each other. This biometric tool can be used to identify an individual. Speech contains the information of an individual like spoken words, speaker's identity, expressions and emotions, accent, living region, health conditions, gender, age, language. A person's authentication and speech as biometric is collectively called Speaker Identification (SI). The main reason of selecting speech biometric trait is due to cheap equipment costs and low time consumption in this work process. Moreover, there is no physical contact or queuing. The objective of this paper is to apply the best algorithms at different stages to provide efficient performance for recognizing speaker in noisy environment in attendance monitoring system.

II. BACKGROUND STUDY

Attendance monitoring in schools and colleges use the conventional method, where the students will be called up one by one or they required to manually signing in the attendance sheet. This process is time consuming, has low accuracy and efficiency. In addition, management have to keep the entire attendance sheet for future references and this is totally a mess if lack of management system in place. Therefore, an easy system is needed to overcome these

problems which can be improved by implement a closed loop automation system.

Traditionally, Fingerprint [4], Thumb impression [5], Hand geometry [6], Iris Recognition [7], Facial Recognition [8] and Voice [9] are used for attendance system. Moreover, a web based attendance system is also proposed which includes GSM/GPRS along with Radio Frequency Identification (RFID) technique in attendance system [10], attendance system using Near Field Communication (NFC) [11]. However, these systems are expensive and have limited use, there are numerous applications today. As all the other biometric traits except speech are difficult to set up due to equipment costs and time consumption in querying process, this section reviews speech as Biometric trait in speaker identification which overcomes the drawbacks of other traits.

There are many two important steps in speaker identification, one is feature extraction and other one is classification. Traditionally, Mel Frequency Cepstral Coefficient (MFCC) is used for feature extraction. In [12], results shows that Gammatone Frequency Cepstral Coefficient (GFCC) algorithm works more efficiently in noisy environment as compared to widely used Mel-Frequency Cepstral Coefficient (MFCC).In [14] when the SNR of test signal changed from 0 to 40 dB, the average accuracy of the MFCCs methods is only 50.05%, while the proposed GFCCs extractors combined to CMN normalization still achieves an average accuracy of 55.43%.

In classification phase of speaker recognition [15] gives that classification plays a crucial part in speaker modeling. The result of classification will strongly affect the speaker recognition engine to decide whether to accept or reject a speaker. Gammatone Mixture model is one of the most well-known models used in speaker identification. In [16] the author works to find the performance of Mel Frequency Cepstral Coefficients (MFCC) in a Gaussian Mixture Model frame work, and compare it to traditional short-time energy and zero-crossing rate feature. It achieved a correct identification close to 90% on MFCC with its first and second derivatives. In [17] different feature normalization techniques are compared. In [18] the author proposed three different popular feature normalization techniques namely MVN (Mean and Variance Normalization), CMN (Cepstral Mean Normalization) and PCA (Principal Component Analysis) and analyzed the result of each technique individually. This paper compared the performance and efficiency of these techniques and evaluates which of these gives the best verification rate. According to result findings CMN gives the best results.

For speech detection there also exist many algorithm like in paper [19] presents different methods of separating voiced and unvoiced segments of a speech signals. These methods are based on short time energy calculation, short time magnitude calculation, and zero crossing rate calculation and on the basis of autocorrelation of different segments of speech signals. From theoretical studies, it has been observed that energy and magnitude for voiced segments is high, whereas ZCR rate is low for voiced signals. Autocorrelation function is used here to show that the voiced segment of speech remains periodic after applying autocorrelation function, while unvoiced signals lose their periodicity. In [20] said ZCR and STM required threshold estimation and proposed new algorithm, also used in this paper for speech detection. It uses Probability Density Function (PDF) of background noise and a Linear Pattern Classifier for extracting voiced portion. Also for de-noising paper [21] uses wavelet filter and additive white Gaussian noise.

This paper is organized with Section III gives the speaker recognition system design, Section IV describes the methodology of work at various stages, and Section V presents the results and performance evaluation.

III. SPEAKER RECOGNITION SYSTEM DESIGN

The process of speaker recognition can be divided into six main stages which are Speech Acquisition, Feature Extraction with post-processing and pre-processing signal feature, normalization and classification as shown in Figure 1.

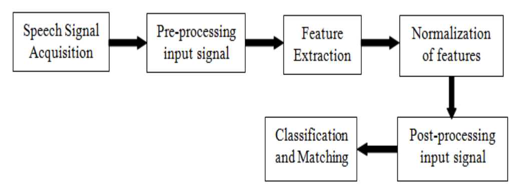


Figure 1. Speaker Recognition System

The initial stage of Signal Acquisition is to replace the speech signal as shown above in Figure 1. The second stage includes pre-processing of input speech signal for extracting speech signal only as signal may contain noise or unvoiced signal. The feature extraction stage is to extract important features of speech signal. The feature post-processing stage is to improve speech results. Next, the Feature Normalization stage is done to avoid the risk of greater influence of greater values in feature vector. Finally, the Classification stage is done for making the input signal with stored speech signal in database.

IV. METHODOLOGY

Speech signal will be captured using the system microphone (e.g. personal computer or laptop). For training process, firstly speech portion is detected from the input signal. Then this speech portion is input to next stage of feature extraction. Feature extraction is done using Gammatone Frequency cepstral coefficient (GFCC) algorithm. Then normalization using Cepstral mean Normalization (CMN) method is done. Further delta features are calculated. Gaussian Mixture Model (GMM) is used for modeling and classification which uses the log likelihood

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

logic to identify the speaker. Figure 2. describes the flowchart of speaker identification design with different techniques used at different stages.

On the other hand, the testing phase starts from utterance of unknown speaker which also go through all the steps same as training phase. But before all these steps, Additive White Gaussian Noise (AWGN) is added and de-noising of signal is done using wavelet filter. This is done to increase the robustness of the speaker identification in noisy environment.

A. *Speech Signal Acquisition*

A Speech signal is captured using the system microphone .It is the cheapest equipment available for acquiring biometric trait. Fig.3 shows acquired signal using laptop microphone. Every signal is analog in nature. To store signal in digital equipment (e.g. computer), there is a need to convert an analog signal into digital signal. As the analog signal is continuous in time and it is required to convert it into digital values. Therefore, sampling rate is defined before acquisition of signal.

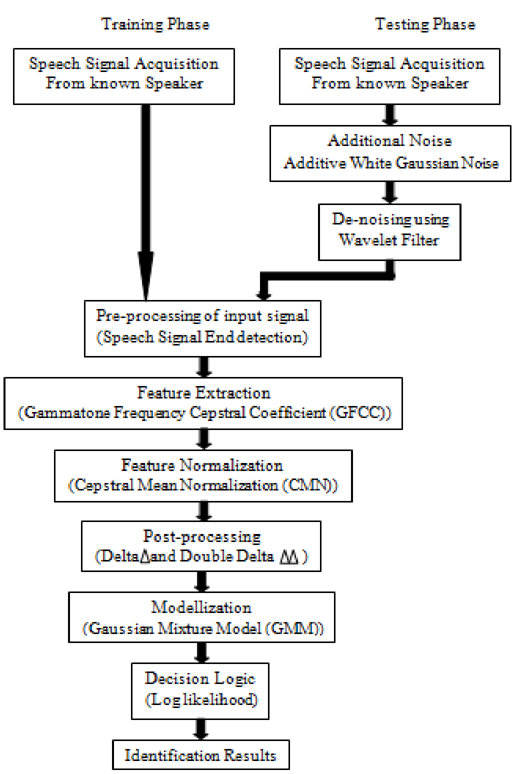


Figure 2. Architecture of Speaker Identification

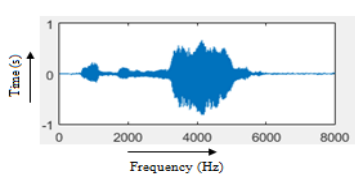


Figure 3. Acquired Signal Using Microphone

B. *Additive White Gaussian Noise(AWGN)*

A noise is not always a useless signal. A noise itself is information that contains information that contains information regarding the source and environment in which it propagates.

White noise is defined as an uncorrelated random noise process with equal power at all frequencies. In communication theory it is assumed that noise is a stationary additive white Gaussian process. Below Figure 4 shows signal before adding noise and Figure 5 shows signal after adding noise.

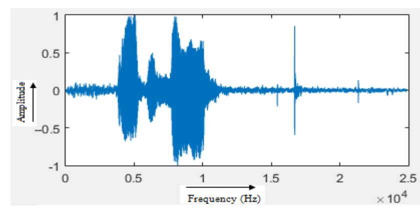


Figure 4. Signal Before Adding White Noise

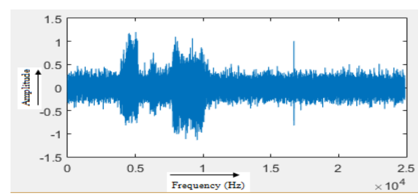


Figure 5. Signal After Adding White Noise

C. *De-Noising with wavelet filter*

Wavelets are used in a variety of fields. In the field of physics wavelets are used for the removal of noise from signals containing information. There are different ways to reduce noise in audio. Wavelets are characterized by scale and position, and are used in analyzing variations in signals and images in terms of scale and position. Because of the fact that the wavelet size can vary, it has advantage over the classical signal processing transformations to simultaneously process time and frequency data. At low scale, compressed wavelets are used. They correspond to fast-changing details, that is, to a high frequency. At high scale, the wavelets are stretched. They correspond to slow changing features, that is, to a low frequency [21]. This paper implements coiflet wavelet for filtering. Following steps shows the filtering process:

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Step 1: *Multilevel decomposition*

For signals, low frequency content shows the identity of signal and high frequency content shows no important information. If high frequency components are removed, still the words are audible and can be recognized easily. However, if enough low frequency components are removed then signal is not clear. Figure 6 shows steps for signal decomposition.

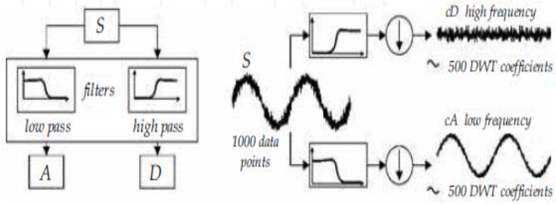


Figure 6. Steps for Signal Decomposition [21].

Decomposition is done in iterative succession. This paper worked on 10 levels of decomposition. Figure 7 shows wavelet decomposition tree.

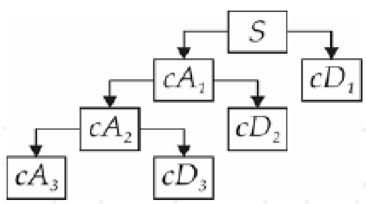


Figure 7. Wavelet Decomposition Tree [21].

Step 2: *Wavelet Thresholding*

Determine noise threshold by using wpbmpen command that returns a global threshold THR for de-noising. THR obtained by a wavelet packet coefficients selection rule by means of using a penalization method provided by Birge-Massart.

Step 3: *Wavelet reconstruction*

Reconstruction is the process of assembling those components back into the original signal without loss of information using threshold values obtained. While being this transformation, it is desirable to establish its investment, i.e. to return to the original signal from the output tree. The mathematical manipulation that affects reconstruction is called the inverse discrete wavelet transforms (IDWT). In order to reconstruct a signal by using Wavelet Toolbox software, reconstruct it from the wavelet coefficients and hard threshold.

D. Pre-processing of input speech

Pre-processing of input signal is to extract only the parts containing speech signal by removing silence or unvoiced portion along with end point detection of speech or voiced portion. This step reduces the dimensions of signals to useful part only which improves the performance and speed of the system. There are a number of algorithms and techniques used like Zero Crossing Rate (ZCR) [19], Short Time Energy (STE) [19]. In this study, Probability Density Function (PDF) of background noise and a Linear Pattern Classifier for extracting voiced portion are used, earlier given in [20]. This algorithm is divided into five steps.

Step 1: Mean (μ) and Standard Deviation (σ) of first 1600 samples (if 8000 is sampling rate) is calculated. Usually, first 1600 samples of speech corresponds silence [20]. Background noise is characterised by Mean (μ) and Standard Deviation (σ). Mean and standard deviation can be analytically written as,

$$\mu = \frac{1}{1600} \sum_{i=1}^{1600} x(i) \qquad (1)$$

$$\sigma = \sqrt{\frac{1}{1600} \sum_{i=1}^{1600} (x(i) - \mu)^2} \qquad (2)$$

Step 2: For each sample of speech, check for one-dimensional Mahalanobis distance function condition.

$$\frac{|x-\mu|}{\sigma} > 3 \qquad (3)$$

where x is an observation, then the sample is voiced, otherwise it is unvoiced or silence. Voiced sample has threshold greater than 99.70 % as in Gaussian distribution and rejects the samples up to 99.70 %. The Gaussian distribution for one-dimension is shown in Figure 8.

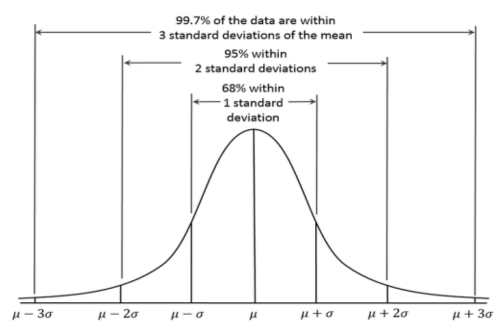


Figure 8. Gaussian distribution

The probability (P) is given as follows:

$$P(\mu - \sigma \le x \le \mu + \sigma) \approx 0.682 \qquad (4)$$

$$P(\mu - 2\sigma \le x \le \mu + 2\sigma) \approx 0.9545 \qquad (5)$$

$$P(\mu - 3\sigma \le x \le \mu + 3\sigma) \approx 0.9975 \qquad (6)$$

Step 3: Mark the voiced sample as 1 and unvoiced as 0. Divide the whole speech into 5 milliseconds (ms) frames.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Step 4: If number of ones in a frame are greater than the number of zeros, denote frame as voiced; otherwise unvoiced.

Step 5: Collect all the voiced frames (labelled 1) and put in a new array.

Figure 9. shows an example of speech input signal and Figure 10. shows voiced portion after implementing the above algorithmic steps.

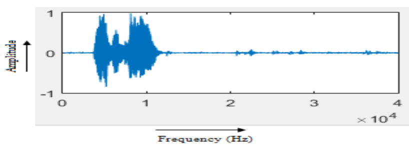


Figure 9. Input Signal

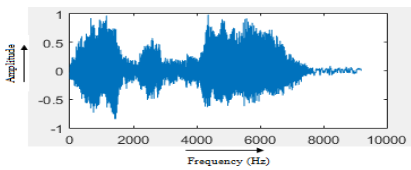


Figure 10. Voiced Signal After Applying Algorithmic Steps

E. *Feature Extraction*

After the detection of voiced signal, feature extraction from signal is done. This stage is to extract those features of speech signal which gives maximum information related to the application. There are many existing algorithms for feature extraction. In this work, Gammatone Frequency Cepstral Coefficient (GFCC) is used which works more efficiently in noisy environment as compared to widely used Mel-Frequency Cepstral Coefficient (MFCC) [12].GFCC is Fast Fourier Transformation (FFT) based technique in speaker identification system.

The detailed process of GFCC extraction is listed as follows [9]:

1. Pass the input signal though a 64 channel gammatone filter bank.

2. At each channel, fully rectify the filter response (i.e. take absolute value) and decimate it to 100 Hz as a way of time windowing.

3. Then take the absolute value afterwards. This creates a Time-Frequency (T-F) representation which is a variant cochleagram.

4. Take the cube root of T-F representation.

5. Apply DCT to derive cepstral features.

In feature extraction stage, firstly a bank of gammatone filter is used for decomposing an input signal into T-F domain. These gammatone filters are standard model of cochlear filtering [12]. 64 filters are used with center frequency range [50, 4000]. Then sampling frequency is decimated to 100 Hz along the time dimension. Further, the magnitudes of the decimated outputs are then loudness-compresses by a cubic root operation. This results into matrix representation T-F decomposition of the input (which is a variant of cochleagram).The cochleagram provides finer frequency resolution at low frequencies than at high frequencies as compares to spectrogram. The following Figure 11. shows cochleagram of an utterance.

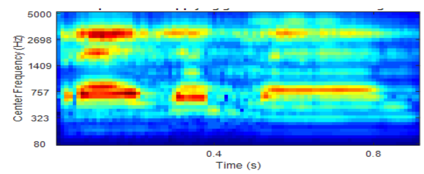


Figure 11. Cochleagram Of Input Speech Signal

Then apply a DCT to the GF featured matrix. Further, GFCC features are extracted using GF featured matrix.

F. *Feature Normalization*

Here, Cepstral Mean Normalization (CMN) is used for normalization. CMN makes it sure that the values of a feature vector have zero mean with one variance. This decreases the risk of larger values that have greater influence on the behavior of different treatments [14].CMN assumes that the mean of Cepstral Coefficients is invariant. Therefore, subtracting the mean will only reduce irrelevant information. Mean and Variance Normalization (MVN) is an extension of CMN which subtracts mean and variance by assuming they are an irrelevant information. Subtract the average of all cepstral coefficients from each cepstral coefficient characterizing the analyzed speech signal. This is called Cepstral Mean Subtraction (CMS).The mean and the variance are given as follows:

$$\mu = \frac{1}{N}\sum_{n=1}^{N} C\,(n) \qquad (7)$$

$$\sigma^2 = \frac{1}{N}\sum_{n=1}^{N}(C\,(n) - \mu)^2 \qquad 8)$$

where, C (n) is a feature vector of n^{th} frame and N is the total number of frames.

$$\hat{C}\,(n) = \frac{C\,(n) - \mu}{\sigma} \qquad (9)$$

where, Ĉ (n) is normalized feature vector.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

The CMN is an alternate way to high-pass filter cepstral coefficients which allow it to compensate the effects of an unknown linear filtering and force the average value of cepstral coefficients.

G. *Feature post processing*

After extracting the GFCC features and double delta features are computed to collect time dynamics. Delta computes the first order derivative of input feature vector and double delta again computes the first derivative of the delta feature vector. Delta, Double delta and GFCC give better results than GFCC feature vector alone. This improves speech preparation result.

H. *Classification*

After extracting the features and removing irrelevant information, there comes classification or modeling or pattern matching. The classification plays a crucial part in speaker modeling. The result of classification will strongly affect the speaker recognition engine to decide whether to accept or reject a speaker. Gammatone Mixture model is one of the most well-known models used in speaker identification.

V. Experimental results

In training phase speech signal is acquired in normal clean environment conditions and then this database of 100 speakers saying phrases like 'Yes mam' "present mam", 'Yes sir', 'present sir' with 4 types of utterances for each phrase with 400 utterances of each is used to generate features using GFCC and then GMM model was used. Then two experiments were conducted in testing phase.

In first phase, white Gaussian noise is added to signal and all steps were performed as in training phase and then at last in decision logic minimum negative log likelihood is used to identify the speaker.

In second phase, white Gaussian noise is added to signal and along this wavelet filter is used for de-noising the signal and then same steps were performed as in testing phase without filter.

Experiment uses the percentage of speaker identification accuracy as a performance evaluation measure for comparing the recognition performances of the feature extractors (i.e. GFCC) used here. Hence, the correct recognition rate (CORR) is adopted for comparison. It is defined as:

$$\%\text{age CORR} = \frac{\text{No.of speech sample correctly classified}}{\text{No.of total samples}} * 100 \qquad (10)$$

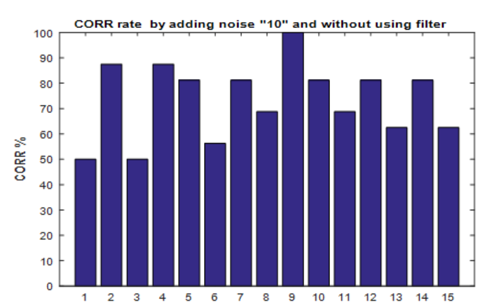
CORR rate has been given person wise in the figures below. Figure 12 shows CORR rate with SNR of 10 without using wavelet filter. Figure 13 shows CORR rate with SNR of 10 with using wavelet filter. Figure 14 shows CORR rate comparison with SNR=10 between GFCC along with wavelet filter and without wavelet filter.

Table 1 show the average identification rate for stored database and real database oriented experiments. In database oriented experiment identification rate is 84.50% when using wavelet filtering and average identification rate is 73.33% when identified without using wavelet filtering and when using real database oriented experiment on an average identification rate is 74% when using wavelet filter and 45% of identification rate when not using wavelet filter.

Results indicates that GFCC algorithm works better with wavelet filter in noisy environment as compared to GFCC algorithm without using wavelet filter in noisy environment. From the graphs it has been found that wavelet filters work well on noisy signals and filter out the added noise easily. In almost all categories of speakers, the wavelet filtering technique better results than without using the filter.

VI. Conclusion

Robust speaker identification with a database of 100 people was collected who pronounced four different phrases of two word pattern. First, 13 GFCC features were extracted using gamma tone filter banks. After that CMN has been applied for normalization of GFCC features. Using extracted GFCC features, first order and second derivative features are also extracted. All three matrices of 13 vectors each are combined and used for speaker identification. Gaussian mixture models are generated from these features using 8 Gaussians. Similar process is repeated for all the entries in the database and 240 total GMM models are generated using 240 samples in the database. After that a sample has been input to white Gaussian noise at 10 SNR and the testing sequence is repeated by using a wavelet filter after adding noise and without using any filter. It has been found that the algorithm gives 85% CORR rate while using wavelet filter and 73% without using the filter. Algorithm is also tried in real time and performance of the real time system gives 74% of identification rate while using filter and 45% of identification rate without using filter. In future the algorithm can be improved for text independent data and real time noisy environment with variation in signal noise ratio.



https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Figure 12. CORR Rate by adding Noise of 10 SNR Without using Wavelet Filter.

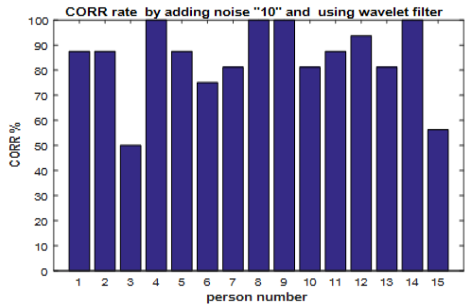


Figure 13. CORR Rate by adding Noise of 10 SNR and using Wavelet Filter.

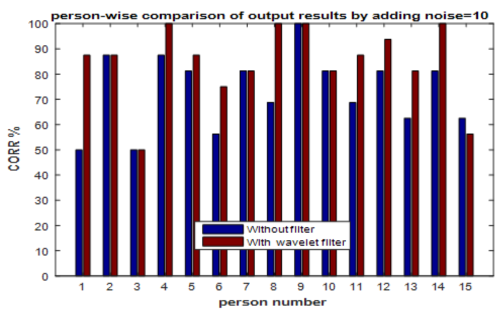


Figure 14. CORR Rate Comparison (when Wavelet Filter is used and when Wavelet Filter not used.)

TABLE 1. PERCENTAGE OF CORRECTLY RECOGNIZED SPEAKERS IN 10 SNR NOISE CORRESPONDING TO FEATURE EXTRACTOR GFCC AND CLASSIFIER GMM ALONG WITH SOUND DETECTION, CMN, ADDITIVE WHITE GAUSSIAN NOISE (AWGN) AND WAVELET FILTER FOR DENOISING.

Database	**Average Correct Recognition rate**		
	Noise (SNR)	***Algorithm without using filter for de-noising***	***Algorithm with wavelet filter for de-noising***
Stored database (100 speakers with 400 utterances each)	10	73.33%	84.50%
Real database (100 speakers with 400 utterances each)	10	45%	74%

REFERENCES

[1] Bhattacharyaa,I., Ghoshb,P., Biswasb,S., “Offline Signature Verification Using Pixel Matching Technique”, International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013. Procedia Technology 10, pp. 970 – 977.

[2] S. Gleni and P. Petratos, “DNA Smart Card for Financial Transactions”, The ACM Student Magazine, (2004), http://www.acm.org.

[3] Soltane, M., Bakhti, M., “Soft Decision Level Fusion Approach to a Combined Behavioral Speech-Signature Biometrics Verification”, in International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 6, No. 1, pp.75-90,February, 2013.

[4] Zainal, N.I., Sidek, K. A., Gunawan,T. S., Mansor, H., Kartiwi, M., “Design and development of portable classroom attendance system based on Arduino and fingerprint Biometric”, In: IEEE International Conference on Information and Communication Technology, pp. 1-4, 2014.

[5] Ujan, I. A., Ismaili. I. A., “Biometric Attendance System”, In: IEEE International Conference on Complex Medical Engineering, pp.499-502, 2011.

[6] Li, T.C., Wu, H.W., and Wu1, T.S., “The study of Biometrics Technology Applied in Attendance Management System” In: IEEE International Conference on Digital Manufacturing & Automation. pp. 943-947. 2012.

[7] Hsiung, T.W., Mohamed, S.S., “Performance of Iris Recognition using Low Resolution Iris Image for Attendance Monitoring”, In: IEEE International Conference on Computer Applications and Industrial Electronics, pp.612-617,2011.

[8] Sajid, M., Hussain, R., Usman, M., “A Conceptual Model for Automated Attendance Marking System Using Facial Recognition” In: IEEE International Conference on Digital Information Management. pp.7-10, 2014.

[9] Dey, S., Barman, S., Bhukya, R. K., Das, R. K., Harish, B. C., Prasanna, S. R. M., Sinha, R., “Speech Biometric Based Attendance System” In: 20th National Conference on Communications, 2014.

[10] Ansari, A. N., Navada, A., Agarwal, S., Patil, S., Sonkamble,B. A., “Automation of Attendance System using RFID, Biometrics, GSM Modem with .Net Framework”, In: IEEE International Conference on Multimedia Technology, pp. 2976-2979,2011.

[11] Balazs Benyo, Balint Sodor, Tibor Doktor, and Gergely Fordo., “Student attendance monitoring at the university using NFC” , In: IEEE, pp. 1-5, 2012.

[12] Zhao, X., Shao Y., Wang D.L., “CASA-based robust speaker identification” IEEE Transactions on Audio, Speech, and Language Processing. vol. 20. pp. 1608-1616, 2012.

[13] Zhao, X., Wang, D. L., “Analyzing noise robustness of MFCC and GFCC features in speaker identification” In: IEEE International conference on acoustics, speech and signal processing, pp. 7204-7208, 2013.

[14] Tazi, E.B., Benabbou, A., Hartl, M., “Efficient Text Independent Speaker Identification Based on GFCC and CMN Methods” In: IEEE International Conference on Multimedia Computing and Systems, pp. 90-95, 2012.

[15] Reynolds, D. A., Rose, R. C., “Robust text-independent speaker identification using Gaussian mixture speaker models” IEEE Transaction Speech and Audio Processing, vol. 3. pp 72-83, 1995.

[16] Jiqing, L., Yuan, D., Jun, H., Xianyu, Z., Haila, W., “Sports audio classification based on MFCC and GMM” In: IEEE International Conference Broadband Network & Multimedia Technology. pp. 482-485 ,2009.

[17] Alam, M.J., Ouellet, P., Kenny, P., O’Shaughnessy, D., “Comparative Evaluation of Feature Normalization Techniques for Speaker Verification”, In: Advance in Nonlinear Speech Process. pp. 246-253, 2011.

[18] Pyrtuh, F., Jelil, S., Kachari, G., Singh, L, J., “Comparative evaluation of feature normalization techniques for voice password based speaker verification” In: IEEE National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, pp. 1-4, 2013.

[19] Jalil, M., Butt, F. A., Malik, A., “Short-Time Energy, Magnitude, Zero Crossing Rate and Autocorrelation Measurement for Discriminating Voiced and Unvoiced segments of Speech Signals”, In: IEEE International Conference on Electronics and Computer Engineering, pp. 20-212,2013.

[20] Saha, G., Chakroborty, S., Senapati, S., “A New Silence Removal and Endpoint Detection Algorithm for Speech and Speaker Recognition Applications”, Department of Electronics and Electrical Communication Engineering Indian Institute of Technology, Kharagpur, Kharagpur-721 302, India.

[21] Villanueva- Luna, A. E., Jaramillo-Nuñez, A., Sanchez-Lucero, D., Ortiz-Lima, C. M., Aguilar-Soto, J. G., Flores-Gil, G., May-Alarcon,M., “De-Noising Audio Signals using MATLAB Wavelets Toolbox”, www.intechopen.com.

[22] Jain, A.K., Ross, A., Prabhakar, S., “An Introduction to Biometric Recognition”, IEEE Transactions on circuits and systems for video technology, vol. 14,no. 1,2004.

[23] Slaney, M., “An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank”, Advanced Technology Group, Apple Computer, pp.1-42,1993.

A novel algorithm for Load Balancing using HBA and ACO in Cloud Computing environment

Seyed Majid Mousavi
University of Debrecen, Faculty of Informatics
Debrecen, Hungary

Fazekas Gábor
University of Debrecen, Faculty of Informatics
Debrecen, Hungary

***Abstract*— Cloud computing is an emerging technology and new trend for computing based on virtualization of resources. Scheduling of tasks to reach load balancing is a challenge in cloud environment. Load balancing is the process of distribution of the load among VMs in order to efficiently utilize of resources and avoiding the situation where some VMs are overloaded or idle. Load balancing of non-preemptive tasks is one of the critical issues in task scheduling in clouds environment. To improve throughput at cloud resources, an intelligent and dynamic load balancing can significantly increase cloud's performance and minimize the costs. Although, many algorithms, strategies and methods have been proposed, but load balancing is still one of the challenging issues in resource allocation in cloud computing environment. In this paper we propose a novel load balancing strategy using Honey Bees and Ant Colony behavior algorithms in cloud environment. The proposed algorithm strives to balance the load of the virtual machines, trying to minimize the completion time of given tasks and reduce response time in cloud infrastructure.**

Keywords: load balancing, ant colony, honey bee, cloud computing.

I. INTRODUCTION

Cloud computing is an emerging internet-based practice to provides computing as a utility service where consumers can pay-per-use. This technology is a collection of thousands of computers interlinked together in a complex manner [1]. The software and hardware resources are allocated to the cloud applications on-demand basis [3].Cloud computing provides a heterogeneous collection of parallel and distributed computing to deliver on-demand access to shared pool of resources. These resources may include a computer, group of computers, network links, central processing units or disk drives [4]. The shared use of resources by the consumers without any strategy brings a range of issues and challenge in cloud environment such as: scalability, fault tolerance, reliability, availability and energy efficiency. These challenges appear when multiple concurrent requests to a single server lead to the server malfunctioning due to overload, while other servers are idle [2]. Nowadays, data centers are mainly heterogeneous; they have physical and virtualized servers from multiple generations and multiple vendors; which means that cloud consumers are geographically dispersed and utilize a diverse range of services. Thus, handling and delivering appropriate services is a major challenge where requests are fluctuating frequently. Since the main objective of load balancing is to reach optimal resource utilization, avoid overloading the system, maximizing throughput and minimize the response time, allocation the available resources with an effective strategy in such a large scale heterogeneous environment is a big challenge.

Load Balancing is a method to distribute workload among resources to make sure that none of existing resources are idle while others are being utilized. If load balancing has been enabled during runtime is called dynamic load balancing. Load balancing in the cloud environment provides opportunities and economies-of-scale, as well as presenting its own unique set of challenges.

Main load balancing goals are [5]: 1) improve the performance substantially 2) maintain the system stability 3) accommodate future modification in the system.

Swarm Intelligence (SI) is defined as the collective problem-solving capabilities of social animals [11]. SI is the direct result of self-organization in which the interactions of lower-level components create a global-level dynamic structure that may be regarded as intelligence [15]. Feedbacks, Randomness and interaction from lower-level components are the main rules of self-organization. Honey bees, ants, flocks of birds and shoals of fish behavior can be a good paradigm for self-organization procedure. Swarm-based optimization algorithms (SOAs) mimic nature's methods to drive a search towards the optimal solution [16].

Honey bee algorithm is a nature inspired Algorithm for self-organization. The performance of this strategy is enhanced with increased system diversity. The major problem in this model is lack of improvement in throughput while the size of system is increased. But, regardless of this issue, when the different kind of service is required, this algorithm is best suited. The ant algorithm is one of the high performance computing methods for combinatorial optimization (CO). Two examples for this strategy is TSP (Travelling salesman problem) and the QAP (Quadratic Assignment Problem).

In this paper, a new load balancing algorithm is proposed, which is a combination of ant colony and honey bee behaviors. This technique aims to develop an intelligent and dynamic load balancing technique in cloud environment to make decisions on its own. In proposed algorithm we used capability of ant colony for combinatorial optimization problem among VMs and performance of honey be algorithm for diverse population of

service types in cloud environment. The contributions of this paper include:

- A new load balancing algorithm by ant colony and honey bee behavior in cloud computing environments.
- Throughput evaluation, performance analysis and comparison of proposed algorithm with other existing algorithms..

II. RELATED WORK

In literature, there are many load balancing techniques and algorithms to increase throughput and efficiency and improve the response time in cloud environment. Each load balancing strategy has its own benefits. [17] [18] [19]. For example Task Scheduling Base (TSB) algorithms in order to better resource utilization achieve load balancing by first mapping tasks to VM and then all VMs to host resources. Opportunistic Load Balancing (OLB) assigns each task in free order to present node of useful .The advantage is quite simple but the completion time (Make span) is very poor. In Round Robin (RR) all the processes are divided and each process is assigned to the processor which load distributions between processors are equal.

In load balancing circumstances, tasks are transferred from over loaded VMs and assigned to under loaded VMs. Load balancing can be categorized base on sender initiated, Receiver Initiated and Symmetric based on who initiated the process, and divided to two categories as: static and dynamic based on the current state of the system [6]. Dynamic load balancing make changes to the distribution of work load among nodes at run-time; they use current load information when making distribution decisions [7]. Thus, dynamic load balancing is executed and shared among all VMs and can affect the overall performance of cloud environments.

Likewise, Sesum-Cavic and Kuhn [20] explicates that dynamic load balancing strategies cover all defects of static load balancing algorithms. Although, to gain dynamic load balancing advantages additional cost for collecting, maintaining and analyzing of load information are required. They have this view that the best approach for load balancing in complex environments are self-organization solutions.

The algorithm proposed in [8] is a Genetic Algorithm. The strategy balances the load of the cloud infrastructure while trying minimizing the makespan of a given tasks set. Simulation results shows that the proposed algorithm is efficient than the existing approaches like Round Robing and Stochastic Hill Climbing (SHC). In [1], proposed an algorithm which aims to achieve well balanced load across VMs for increase the throughput. The proposed algorithm also balances the priorities of tasks on the machines in such a way that the amount of waiting time of the tasks in the queue is minimal.

Shang-Liang et al. [9] consider both processing ability and loading. They propose a new strategy for load balancing, which can be used to both virtual and physical environment. Jung et al. [13] propose an algorithm for location-aware dynamic resource allocation. And also a comprehensive comparison among different resource allocation strategies is covered in [10]. In proposed Hua's paper [12] algorithm which is based on ant colony optimization for resource allocation, all the specifications in cloud environment are considered. In compare with genetic and annealing algorithm, proposed algorithm has more efficiency and proved that have more throughputs in cloud and distributed environments. The paper in [17] presents a scheduler model based on ACO to allocate VMs to physical Cloud resources. The aim of the model is using ACO capability for scheduling in an online Cloud scenario while multiple users connect to the Cloud at different times to execute their parameter sweep experiments (PSEs).

III. METHODOLOGY

Self-aggregation is a process of attraction and repulsion for load balancing technique that classify services together based on specific characteristics. The proposed system performs load balancing by pre-emptive scheduling. Loading of a task to VMs and removed tasks from over loaded VMs are similar to a honey bee foraging. Let assume a set of m Virtual machines as resources and a set of n tasks. Total length of tasks which is allocated to a virtual machine *i* is called the load of VM*i*. If total tasks which are allocated to VM*i* as load on a single virtual machine is equal to number of tasks at the time *t* on VM*i*'s queue on service rate of the VM*i* at time t, then load VM*i* is:

$$L_{VM_i} = N(T,t) / S(VM_i, t)$$

and load of all VMs are:

$$L = \sum_{i=1}^{n} L_{VM_i}$$

The maximum load $L_{\max} = \max_{i \in VM} L_i$ is called makespan or CT_{max}. Makespan is maximum completion time of all tasks and is defined as the following function.

$$C_{\max} = Max(Complition\_Time[i, j])$$

$$\{i \in \text{Task} \mid 1 \le i \le \text{N}, j \in \text{VM} \mid 1 \le j \le M\}$$

Response time is defined as interval time between submission a task in a VM and the first response that is produced. Load balancing has direct effective on reduction of response time and improving responsiveness of the VMs. In order to reduce makespan and response time, tasks will be transferred and assigned from one busy VM to an idle VM during load balancing process.

To balance the loads among VMs, each VM announce its capacity C_{VM_i} based on number processors in VM*i*, MIPS of each processor, communication bandwidth ability of VM*i*, to other VMs in cloud environment.

Overall capacity of all VMs is:

$$C = \sum_{i=1}^{n} C_{VM_i}$$

VM*i* receives this information from VM*i* with a finite delay τ_{ij} . Each VM*i* then use this information to estimate of the average number of tasks in the capacities of the *n* VMs in cloud environment. Thus, a simple estimation is:

$$C_{avg} = \sum_{i=1}^{n} C_{VM_i} \; / \; n$$

VM$_i$ then compares its capacity, with its estimate of the C_{avg} .

If $C_{VM_i} \leq C_{avg}$ then VM*i* transfer task to the other VMs else accept task.

The pheromone update process to transfer rejected task from VM*i* to VM*j* determine the new pheromone value, as follows:

$$\tau_{ij} \leftarrow (1-\rho).\mathrm{C}_{avg} \sum\nolimits_{k=1}^{m} \Delta_{ij}^{k}$$

The probability which task *k* currently at VM*i* choosing to go to VM*j* is:

$$P_{ij}^{k}(t) = \frac{[\tau_{ij}(t)]^{\alpha}[\eta_{ij}(t)]^{\beta}}{\sum\nolimits_{k=1}^{n} \tau_{ij}^{\alpha}.\eta_{ij}^{\beta}}$$

where

$$\begin{cases} \tau_{ij} = \text{Pheromone trail} \\ \eta_{ij} = \text{Heuristic value} \\ \alpha, \beta = \text{Reative influence of the pheromone trail parameters} \end{cases}$$

Processing time for VM*i* is $P_{T_i} = L_{VM_i} / C_i$, and thus, $P_T = L / C$.

In this circumstances standard deviation of load is: $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_{T_i} - P_T)^2} + P_{ij}^{k}(t)$.

If the standard deviation of the VM load is under or equal to the threshold condition set (Ts) [0–1] then the system is balanced [13].

<u>The proposed Algorithm is defined as follow:</u>

Step 1: Calculation of L
    If $\sigma <= Ts$
        System is balanced
    ElseIf $L > C_{max}$
        Load balancing is impracticable.
Step 2: Task Scheduler Schedules job to different VMs
    While (! balancing):
    Calculation of C, C_{avg}, C_{VM_i}
    If $C_{VM_i} > C_{avg}$
        Accept task$_i$
Step 3: Task seeks its surrounding VMs for availability with probability

$$P_{ij}^{k}(t) = \frac{[\tau_{ij}(t)]^{\alpha}[\eta_{ij}(t)]^{\beta}}{\sum\nolimits_{k=1}^{n} \tau_{ij}^{\alpha}.\eta_{ij}^{\beta}}$$

Step 4: If VM is available then
        Transfer task$_i$ to VM$_j$
    Else
        Go to Step 3
    EndWile

IV. EXPERIMENT RESULTS AND DISCUSSIONS

Evaluation of efficiency and performance in distribute environments based on effectiveness of workload models under different system and configurations and requirement is very difficult. Cloudsim is one of the best toolkits to implement provisioning techniques and can be extended easily with limited effort. It is an extensible simulation toolkit that enables modeling and simulation of Cloud computing systems and application provisioning environment [14]. In order to validate the efficiency of the proposed algorithm, several scenarios are examined. In first experimentation, as it has been described in Table 1 the execution status of 8 tasks has been shown which are running in three VMs with different load. The first three tasks are allowed to run in VM0, VM1 and VM2 according to proposed algorithm. Task 3, 4, 5 are also allocated to VM0, VM1 and VM2 based on their load. Task 6, 7 are transferred to VM1 and VM2 because VM0 is in overload.

Cloudlet ID	Status	Datacenter ID	VM ID	Time	Start time	Finish Time
0	SUCCESS	2	0	1200	0	1600
1	SUCCESS	2	1	2000	0.1	1400.1
2	SUCCESS	2	2	400	200.1	600.1
3	SUCCESS	2	0	1500	1600.1	3100.1
4	SUCCESS	2	1	600	1400.1	2000.1
5	SUCCESS	2	2	800	1300.1	2100.1
6	SUCCESS	2	1	1100	800.1	1900.1
7	SUCCESS	2	2	1720	400.1	2120.1

Table 1: proposed algorithm's execution status in cloudsim simulation

We also have evaluated and analyzed the performance of proposed algorithm based on the result of Cloudsim simulator. Figure 1depicts comparison of completion time before and after load balancing by proposed algorithm. Figure 1 a and b: illustrate response time difference in both load balancing and unbalanced situation.

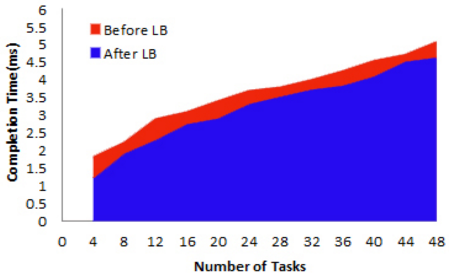


Figure 1 a): Depicts Makespan before and after load balancing

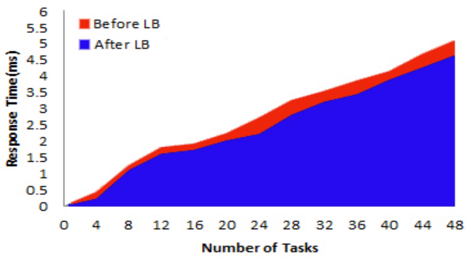


Figure 1 b): Depicts Response Time before and after load balancing

In next experimentation, we configured the simulation parameters according to one region (table 3) and two regions (table 4) around the world (table1). in the first scenario all users connect to one data centers with having 5, 10, 15 VMs. the simulation result are given in Table 3 with calculated average Completion Time for proposed algorithm, Round Robin, Random and Weighted least connections.

S.No	*User Base*	*Region*	*Online-users during peak hrs.*	*Online -users during off-peak hrs.*
1	UB1	0- N.America	5,000,000	110,000
2	UB2	1- Europe	4,200,000	200,000

Table 2 : simulation parameters for two different regions

S.NO	*DC*	*CT for PP*	*CT for R*	*CT for RR*	*CT WRR*
1	With 10 VMs	227.33	229.21	231.32	230.99
2	With 15 VMs	202.43	227.25	231.01	230.49
3	With 20 VMs	198.21	220.12	231.44	229.80

Table 3: Simulation scenario and calculated average Response in one DC in (ms)

Figure 2: comparison of calculated average Response Time

	DC	*Group*	*CT for PP*	*CT for R*	*CT for RR*	*CT WRR*
1	With 10 VMs each DC	T1	275.32	280.41	282.23	291.32
2	With 15 VMs each DC	T2	220.21	277.45	270.54	290.03
3	With 20 VMs each DC	T3	219.11	260.32	250.05	270.39
4	(10VMs DC1) (15VMs DC2)	T4	210.03	240.87	240.32	260.66
5	(10VMs DC1) (20VMs DC2)	T5	210.88	239.03	248.21	255.54
6	(15VMs DC1) (20VMs DC2)	T6	216.10	240.53	247.93	253.54

Table 4: Simulation scenario and calculated average Response in two DC in (ms)

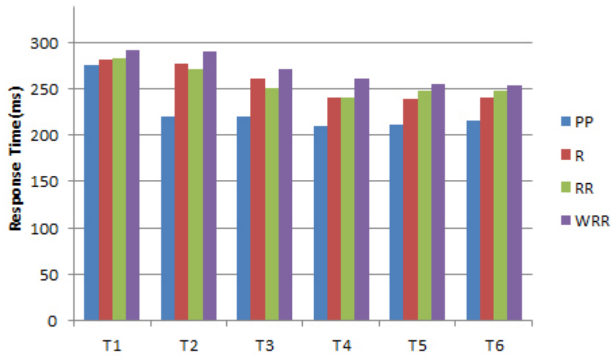


Figure 3: comparison of calculated average Response Time

V. CONCLUSION

Although there are many solutions to improve the performance of load balancing in cloud environment but present new solutions would be effective to increase current efficiency and throughput, even if there is minor improvements and performance. In this paper, a new cloud load balancing algorithm is proposed by comparing previous studies. The proposed new paradigm can be applied to virtualized cloud environments. This algorithm also can consider the priority of tasks. Proposed strategy for load balancing improves throughput of cloud environment and focuses on reducing of completion time and also waiting time of a task in queue of the VM. Therefore, the algorithm reduces response time of VMs. This load balancing strategy is appropriate for heterogeneous systems and non-preemptive tasks in cloud environments.

REFERENCES

[1] Dhinesh Babu L.D. and P. Venkata Krishna. Honey bee behavior inspired load balancing of tasks in cloud computing environments. Applied Soft Computing, 13(5):2292 { 2303, 2013. ISSN 1568-4946.

[2] Jing Yao and He Ju-Hou. Load balancing strategy of cloud computing based on arti_cial bee algorithm. In Computing Technology and Information Management (ICCM), 2012 8th International Conference on, volume 1, pages 185{189, April 2012.

[3] V. Vinothina, Dr. R. Shridaran, and Dr. Padmavathi Ganpathi, A survey on resource allocation strategies in cloud computing, International Journal of Advanced Computer Science and Applications, 3(6):97--104, 2012.

[4] Z. Bo, G. Ji and A. Jieqing, “Cloud loading balance algorithm,” Proceedings of IEEE 2nd International Conference on Information Science and Engineering (ICISE), Hangzhou, China, pp. 5001–5004, 2010,.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

[5] David Escalante and Andrew J. Korty, Cloud Services: Policy and Assessment, EDUCAUSE Review,vol. 46, no. 4 (July/August 2011)

[6] Ali M. Alakeel, A Guide to Dynamic Load Balancing in Distributed Computer Systems, IJCSNS International Journal of Computer Science and Network Security,VOL.10 No.6, June 2010.

[7] N. Malarvizhi, V. Rhymend Uthariaraj, Hierarchical load balancing scheme for computational intensive jobs in Grid computing environment, in: Advanced Computing, 2009. ICAC 2009. First International Conference on, 13–15 Dec, 2009, pp. 97–104.

[8] Kousik Dasgupta, Brototi Mandal, Paramartha Dutta, Jyotsna Kumar Mandal, Santanu Dam, A Genetic Algorithm (GA) based Load Balancing Strategy for Cloud Computing, Procedia Technology, Volume 10, 2013, Pages 340-347, ISSN 2212-0173.

[9] Shang-Liang Chen, Yun-Yao Chen, Suang-Hong Kuo, CLB: A novel load balancing architecture and algorithm for cloud services, Computers & Electrical Engineering, Available online 28 February 2016, ISSN 0045-7906,

[10] Chandrashekhar S. Pawar and R.B. Wagh, A review of resource allocation policies in cloud computing, World Journal of Science and Technology, 2(3):165-167, 2012.

[11] Bonabeau, E.; Dorigo, M.; Theraulaz, G. Swarm Intelligence: From Natural to Artificial Systems; Oxford University Press: New York, NY, USA, 1999.

[12] Hua Xiayu, Zheng Jun, Hu Wenxin. Ant colony optimization algorithm for computing resource allocation based on cloud computing environment [J].Journal of East China Normal University (Natural Science). 2010.

[13] Gihun Jung and Kwang Mong Sim, Location-Aware Dynamic Resource Allocation Model for Cloud Computing Environment, International Conference on Information and Computer Applications(ICICA), IACSIT Press, Singapore, 2012.

[14] Calheiros, Rodrigo N., et al. "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms." Software: Practice and Experience 41.1 (2011): 23-50.

[15] Koc, E. The Bees Algorithm Theory, Improvements and Applications. Ph.D Thesis, Cardiff University, Cardiff, UK, 2010.

[16] Yuce, Baris, et al. "Honey bees inspired optimization method: the bees algorithm." Insects 4.4 (2013): 646-662.

[17] S.-C.Wang, K.-Q. Yan, S.-S.Wang, C.-W. Chen, “A three-phases scheduling in a hierarchical cloud computing network”, in: Communications and Mobile Computing (CMC), 2011 Third International Conference on,IEEE, 2011, pp. 114–117.

[18] O. Elzeki, M. Reshad, M. Elsoud, “Improved max-min algorithm in cloud computing, International Journal of Computer Applications”vol 50 (12) (2012)pages 22–27.

[19] U. Bhoi, P. N. Ramanuj, “ Enhanced max-min task scheduling algorithm in cloud computing”, International Journal of Application or Innovation in Engineering and Management (IJAIEM), ISSN,2013,pages 2319—4847.

[20] Vesna A Sesum-Cavic and Eva Kuhn. Chapter 8 self-organized load balancing through swarm intel-ligence. In Nik Bessis and Fatos Xhafa, editors, Next Generation Data Technologies for Collective Computational Intelligence, volume 352 of Studies in Computational Intelligence, pages 195{224. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-20343-5.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Route Optimization in MANET Using Hopfield Neural Networks: MANET-HOP

Sanjeev Gangwar
Department of Computer Application
V. B. S. Purvanchal University
Jaunpur, India

Dr. Krishan Kumar
Deprtment of Computer Science
Gurukul Kangri University
Haridwar, India

***Abstract*—As we know that Mobile Ad Hoc Network is the combination of nodes having unstable setup which usually formed instantly in independent manner. It does not have any centralized administration. Moreover they don't have any permanent setup and routers. In such situations routing becomes the responsibility of individual nodes and also routing is equally important to realize the practical benefits of MANET. Traditional protocols of MANET: DSR, AODV, DSDV, OLTP work well but still need improvements time-to-time as per the new issues like QoS provisioning and routing. Above protocols mainly depends on hop count measurement. In this paper we have implemented a specific problem of six nodes situated at different locations with primary goal to find the shortest route visiting each node at least once which is based on the concept of Travelling Salesman Problem using Feedback/Hopfield Neural Network. And we found that Hopfield networks are suitable to find the shortest route.**

Keywords- Mobile ad-hoc network, Hopfield neural network, Travelling salesman problem, Route optimization

I. INTRODUCTION

Mobile Ad-hoc network or simply MANET is an infrastructure-less, dynamic network consisting of wireless mobile nodes that can communicate among each other without any centralized control. It is a system of mobile nodes (laptops, sensors, etc.) interfacing without the assistance of centralized infrastructure (access points, bridges, etc.) Therefore many kinds of security attacks like worm hole attack, rushing attack which may harm to MANET because of their basic characteristics e.g. Distributed operation, wireless medium, dynamic topology etc. If we compare MANET with wired networks they show some special features like dynamic topology, limited bandwidth, and unstable shared wireless media. Such dynamic features give birth to serious issues and challenges to routing protocols which enhance the adaptability to the dynamic environment.

Due to increase in the demand of the computers in our daily life, it increases the demand of connectivity. Through connectivity of various nodes in the network, these nodes in the network can easily share their data or objects. Wired network have been used for a long time. Due to some restrictions of wired network, requirement for the wireless network has been increased for sending messages, emails and communicate with other. So MANET have been developed which comprises of a large number of nodes. In mobile ad-hoc network, nodes can communicates with the other nodes without any need of central administration or base station. MANET is commonly used for all purpose like offices for doing work and colleges for maintain details [2].

It normally has a rapid changing network topology in which the nodes roam from here to there i.e. they are highly mobile devices. This behaviour requires routing protocols that dynamically discover routes rather than conventional distance vector routing protocols. Another very important issue that is IP sub-netting is also not possible because MANETs are highly dynamic, hence this must be resolved. Then there is power depletion of nodes due to large number of message passed during cluster formation and limitation of battery power. Links in MANET are not symmetric at all times. If a routing protocol is dependant only on the bi-directional links, the connectivity and size of the network may be restricted severely. A protocol that makes use of unidirectional links as well as bidirectional link scan significantly lessens the network partitions and enhances routing performance. Normally three types of routing protocols for MANETs are studied [4],[5].

A. *Proactive Protocol*

These are the protocols which maintain their database by exchanging information among their neighbouring nodes in network regularly in a proactive manner hence called proactive protocols. It also tries to maintain the information consistency and routing tables up to date in the network. This is done by propagating, proactively, route updates at fixed intervals. As the resulting information is normally maintained in tables, the protocols are sometimes also referred to as table-driven protocols.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

B. Reactive Protocol

Those protocols which do not need to maintain their path information in advance but maintain as per the requirement dynamically from source to destination. After discovering the path/route information the entire responsibility is of node until the route is no longer used or has expired. And also, until the destination becomes accessible.

C. Hybrid Routing Protocol (CBRP)

This is the combination of both proactive and reactive protocols. The hybrid protocols makes use of best features of both proactive and reactive routing which helps to overcome the frequently changing topology problem in MANET. Different types of Pro-active protocols are given in table-1.

TABLE I. DIFFERENT MANET PROTOCOLS

S.No.	Proactive Protocols	Reactive protocols	Hybrid protocols
1	Destination Sequenced Distance Vector (DSDV)	Ad hoc On Demand Distance Vector (AODV)	Zone routing protocol (ZRP)
2	Optimized Link State Routing (OLSR)	Dynamic Source Routing (DSR)	Greedy Parameter Stateless Routing (GPSR)
3	Cluster head Gateway Switch Routing (CGSR)	Associativity Based Routing (ABR)	
4	Fisheye State Routing (FSR)	Temporally Ordered Routing Algorithm (TORA)	
5	Routing Protocol (WRP)		

Besides the availability of other technologies Mobile Ad hoc network is still good choice of network designers due to security and cost effectiveness in different types of application areas like military, tracking, and environment etc. Besides many features, MANET has many issues like battery power, routing etc. Hence, to handle these problems, we being the researcher often work on routing different techniques e.g. DSR, AODV, DSDV, and OLTP. These techniques are effective, efficient, and power saving during the time of data transfer. Table 1 and figure 1 collectively show hierarchy of theses routing protocols.

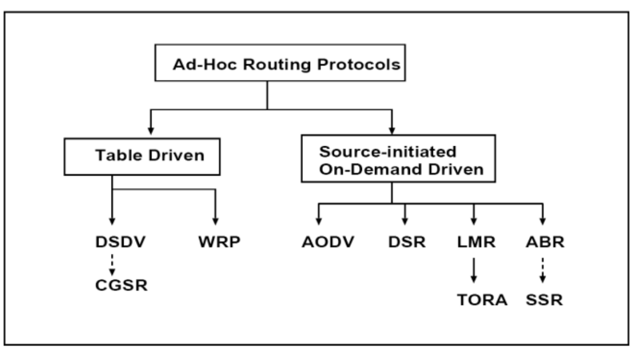


Figure 1. The Family tree of Ad-Hoc Rouitng protocols

Today, artificial neural network (ANN), a widely used & popular technique of soft computing, has solved most promising real time problems for many real time applications like function approximation, weather forecasting, stock exchange prediction, air traffic control etc. It is also best suited to find out the path between nodes. Generally two popular neural networks methods: self organizing map (SOM) and Hopfield are used to find out the path among nodes. Or we can say that all problems related with graph theory can be solved using these two methods.

In this paper, we have used the Travelling Salesman Problem (TSP), which is NP complete problem as it does not have any algorithm. Hopfield method, a recurrent or feedback network has been implemented to find out the near optimal path based on the concept of from source to destination with some constraints.

II. PROBLEM DESCRIPTION

The problem has been taken from the particular scenario of MANET. Our problem is very simple in which we have considered six nodes at different locations distances from each other as shown in figure-1. Theses nodes have some parametric values also which are given in table-1. These parametric values include speed of nodes, energy level etc.

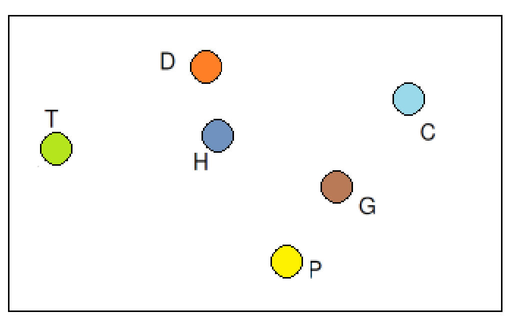
Also from the definition of an undirected graph, the underlying topology of the MANET can be generalized as G = (V,E), where V is the set of vertices (N nodes), and E is the set of its edges. Further, a link cost matrix L = Lij, where Lij is the cost from node i to node j, s is the source node and d is the destination node. For each link (i, j), there exists a nonnegative number Lij, called the cost including the time delay, the bandwidth, and the traffic load of the link from node i to node j. If there is no link from node i to j, Lij is set to a very large value in order to exclude it from the routing path. Note that link (i, j) is symmetric with link (i, j) and Lij=Lji since the network is given by an undirected graph. If we define an undirected path Psd for a routing problem, an ordered sequence of nodes connecting s to d can be written by: P^{sd} = (s,a,b, ...i,d), where the route can be give as

$$(s \rightarrow a \rightarrow b \rightarrow c \cdots \rightarrow i \rightarrow d)$$

In our case source and destination both are T. And other nodes are D,H,P,G,C. In this case we can also find out the total cost on the basis of shortest path, and if we are able to find the total cost for the shortest path then we can easily find out the total minimum cost of the path. Which can be given as

$$TC^{sd} = L^{sa} + L^{ab} + \cdots + L^{id}$$

In our case we are not going to discuss it.



https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Figure 1. Six nodes positions

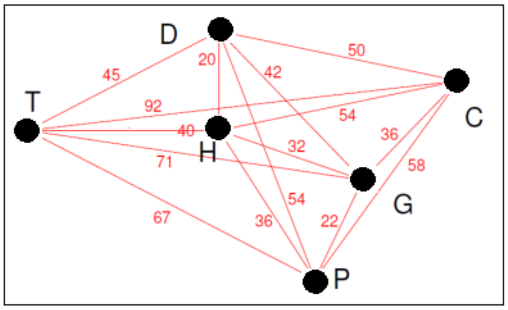


Figure 2. Position & respective Distances of the six nodes from each other

TABLE II. DISTANCE BETWEEN EACH PAIR OF LOCATIONS IN METERS

	1	2	3	4	5	6
1	0	36	32	54	20	40
2	36	0	22	58	54	67
3	32	22	0	36	42	71
4	54	58	36	0	50	92
5	20	54	42	50	0	45
6	40	67	71	92	45	0

TABLE III. PARAMETERS USED

Parameter	Values
Number of nodes	6
Maximum speed	25 m/s
Minimum speed	0 m/s
Node flows	6
Simulation time	15 s
Packet size	512
Traffic type	CBR
Dimension of space	100 x 100
Initial node energy	600 w
Power consumption Pr	0.1 w
Power consumption Pt	1.0 w
Power consumption piddle	0.5 w

Our main goal is to find out the Hamiltonian circuit or shortest distance starting from cluster head and visiting all the nodes at least once. The problem is similar to the Travelling Salesman Problem (TSP) of algorithms. TSP is not a specific problem but it is a concept which can be applied in any topological network where we need and focus to find out the shortest distance. We have used Hopfield network or simply Hopnet to optimize the path. We can also use other methods to find the same eg Dikshatra algorithm, DFS, BFS etc.

III. RELATED WORK

I have proposed an implementation of an algorithm of artificial neural network to find the approximate solution for Travelling Salesman Problem. TSP is a good example of constraint satisfaction and optimization and hence belongs to the NP complete problem. So far many researchers have done this type of work by using different methods like Dikhjatra Algorithm, Branch and Bound method, Breadth First Search (BFS), Depth First Search (DFS), Hamilton circuit etc. We have used Continuous Hope field neural network to find the solution for the given problem. The algorithm gives near optimal result for six nodes in our case.

In [2], an on-demand routing protocol, UBPCR [utility-based power control routing] has been used which normally reduces the trade-offs that arise in the other energy-aware route selection mechanisms. The approach is basically based on the rule of economic framework that gives the value of link's satisfaction. In this, during the route-searching, each intermediate node is executed via two different consecutive phases: the scheduling phase and the transmit power control phase.

In [3], MPR selection is defined, as we know that MPR selection is very important and critical function. And hence link state routing (OLSR) protocol is proposed. This paper proposes a Fuzzy logic based novel routing metric for MPR selection based on the energy, stability and buffer occupancy of the nodes. An algorithm is designed to cope with these constraints in order to find quality MPR (QMPR) that guarantees the QoS in OLSR. The aim of this paper is to formulate, build, evaluate, validate and compare rules for QMPR selection using fuzzy logic [3].

Infrastructures less networks have no fixed routers; all nodes are capable of movement and can be connected dynamically in an arbitrary manner. Nodes of these networks function as routers which discover and maintain routes to other nodes in the network. Topological changes in mobile ad hoc networks frequently render routing paths unusable. Such recurrent path failures have detrimental effects on quality of service. A suitable technique for eliminating this problem is to use multiple backup paths between the source and the destination in the network. Most proposed on-demand routing protocols however, build and rely on single route for each data session. Prediction is done by using a Multi-Layer Perceptron (MLP) Network which is trained with back propagation error algorithm. Experimental results shows the MLP net can be a good choice to predict the reliability of the links between the mobile nodes with more accuracy [6].

Evolutionary algorithms or simply EA is also very popular method now-a-days for finding the output when we have large amount of data. By the genetic structures and the genetic operators for generating new variants, evolutionary algorithms can be classified as GA, evolution strategies (ES), evolutionary programming (EP) and genetic programming (GP). GA is often able to automatically acquire and accumulate implicit knowledge about the search space during its search process and self-adaptively control the search process through a random optimisation technique. When applied to TSP, the GA based routing algorithm takes a few seconds to run the 100 generations for a neural network [7],[8].

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

In [13], C. W. Ahn et al. were proposed a optimal routing algorithm using Hopfield neural network (HNN). In the proposed work, the neurons use whole information which was available and also correlated the information with local neuron. By this method, they obtained a faster convergence. Consequently, a better route is optimized algorithms has been emerged based on HNN.

In [14], A. W. Mohammed et al. have proposed a PSO-based search algorithm. In the proposed algorithm path-encoding scheme is used which based on priorities. During path creation process, a heuristic operator has been used for reducing invalid loop creations. In comparison to other GAs algorithms it was claimed that the PSO-based SP algorithm is better one.

In [15], C. Perkins et al. have proposed an AODV routing protocol. AODV is most popular routing protocol used on-demand routing protocols in MANET. A source host broadcasts a RREQ packet when it needs a route to a specific destination host. Each receiving host that receives the RREQ packet checks whether it is the destination, if it is the destination, then it sends a RREP packet. If it not destination, then it rebroadcasts the RREQ packet via intermediate hosts. In AODV the established route has no knowledge about the network status.

In [16], H. XIAO et al. have proposed the FQMM as the first Quality of Services model for MANET which is a hybrid of both Integrated Service and Differentiated Services architectures. The proposed method FQMM includes many features such as adaptive conditioning, dynamic roles of nodes and hybrid provisioning.

In [17], P. K. Suri et al. have proposed a routing technique named as Cluster based QoS routing (CBQR). The main focus of this protocol is on bandwidth efficiency. It is a table driven routing protocol which deals with the bandwidth requirement over the wireless network. This protocol also takes care of the stale routes, storage overheads and limited battery power.

In [18], S. -J. Lee et al. have proposed a Dynamic Load Aware Routing (DLAR) protocol. It defines the network load of a mobile node as the number of packets in its interface queue.

In [19], H. Hassanein et al. have proposed a Load-Balanced Ad hoc Routing (LBAR) protocol. It defines the network load in a node as the total number of routes passing through the node and its neighbours.

In [20], K. Wu et al. have proposed a Load-Sensitive Routing (LSR) protocol. In proposed protocol, network load in a node is defined as the summation of the number of packets being queued in the interface of the mobile host and its neighboring hosts. LSR is more accurate than among these protocols of DLAR or LBAR on the basis of load metric. LSR protocol generates the contention delay that increases the overall delay of transmission.

In [21]. S-T. Sheu et al. have proposed a Delay-Oriented Shortest Path Routing protocol. In IEEE 802.11 wireless networks, the contention delay problem is minimized in this protocol by analyzing the medium access delay of a mobile node.

IV. Methodology Used

To find out the minimum route in a MANET is a very important and common issue which can be find out by using various methods. Travelling Salesman Problem (TSP) is the well method known in optimization. The TSP problem is NP-complete problem. There is no approximation algorithm for this problem, which could give a perfect solution. Thus any algorithm for this problem is going to be impractical with certain examples. Here we assume that we are given n cities, and a non-negative integer distance D_{ij} between any two cities i and j. We try to find the tour or path for the salesman that best fits the mentioned criterion. There are various neural network algorithms that can be tried to solve such constrain satisfaction problems. Most solution have used one of the methods: Hopfield Network, Kohonen Self-organizing map, Genetic Algorithm.

Here an approximate solution is found for TSP using Hopfield neural network (HNN). HNN, is a good candidate for implementing the shortest path computations involved in the routing problem, primarily owing to the potential of the neural network hardware approach for high speed computation. We will use six nodes and hence 36 neurons in all. And draw first image showing the six nodes as per the distances from each other. These shall be shown in the table.

A. Travelling Salesman Problem

Travelling Salesman Problem or TSP is a very famous concept of design and analysis of algorithm which is used to find out the shortest path from the desired node. As per TSP there is a list of cities that are to be visited by a salesman. A salesman starts from a city and come back to the same city after visiting all the cities. Here the objective is to find the path, which follows following constrains:

1) *Step-1:* Salesman has to visit each city. He should not leave any city unvisited.
2) *Step-2:* Each city should be visited only once.
3) *Step-3:* The distance that he travels till he returns back to the city he has started should be minimum.

The TSP can be described as follows:
TSP = {(G, f, t): G = (V, E) a complete graph,
f is a function V×V Z, → t ∈ Z,
G is a graph that contains a travelling salesman tour with cost that does not exceed t}. See figure-3

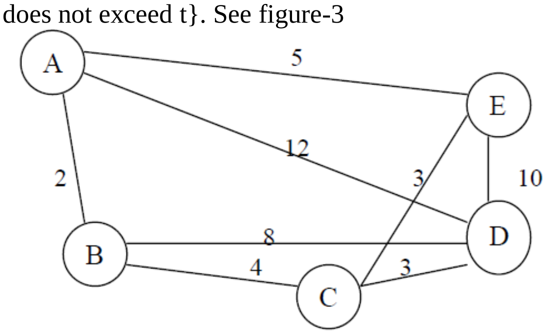


Figure 3. A graph with weights on its edges

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

The problem lies in finding a minimal path passing from all vertices once. For example the path Path1 {A, B, C, D, E, A} and the path Path2 {A, B, C, E, D, A} pass all the vertices but Path1 has a total length of 24 and Path2 has a total length of 31.

B. Travelling Salesman Problem is NP-complete

First, we have to prove that TSP belongs to NP. If we want to check a tour for credibility, we check that the tour contains each vertex once. Then we sum the total cost of the edges and finally we check whether the cost is minimum or not. This can be completed in polynomial time thus TSP belongs to NP.

Secondly we prove that TSP is NP-hard. One way to prove this is to show that Hamiltonian cycle ≤ TSP (given that the Hamiltonian cycle problem is NP-complete). Assume G = (V, E) to be an instance of Hamiltonian cycle. An instance of TSP is then constructed. We create the complete graph G′ = (V, E′), where E′ = {(i, j): i, j ∈ V and i ≠ j. Thus, the cost function is defined as:

$$t(i\,,j\,) = \begin{cases} 0 \, if\ (i,j) \in E \\ 1 \, if\ (i,j) \notin E \end{cases}$$

Now suppose that a Hamiltonian cycle h exists in G. It is clear that the cost of each edge in h is 0 in Gas each edge belongs to E. Therefore, h has a cost of 0 in G′′. Thus, if graph G has a Hamiltonian cycle then graph G has a tour of 0 cost. ′Conversely, we assume that G’ has a tour h’ of cost at most 0. The cost of edges in E’ are 0 and 1 by definition. So each edge must have a cost of 0 as the cost of h’ is 0.

We conclude that h’ contains only edges in E.
So we have proven that G has a Hamiltonian cycle if and only if G’ has a tour of cost at most 0. Thus TSP is NP-complete.

C. Hopfiel Neural Networks or HNN

Application to the constrained combinatorial optimisation problems using artificial neural networks (ANN) was first introduced by Hopfield and Tank in an attempt to find good solutions if not the best solution within a permissible time period by applying their model, the HNN, to the travelling salesman problem (TSP) [9]. The convergence of the nonlinear dynamic system for symmetric connections was verified by introducing the Lyapunov energy function. Since then, the HNN has been successfully used to solve various optimisation problems known as NP-complete [10].HNN is a dynamic network, which iterates to converge from an arbitrary input state. It works as minimizing an energy function and fully connected network like mesh topology. It is a weighted network where the output of the network is fed back and there are weights to each of this link. The main task is set the number of neurons and value of weights.

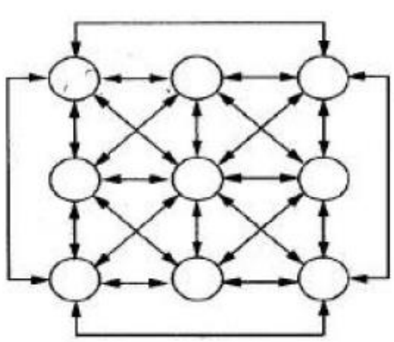


Figure 4: Fully Connected Hopfield Network for TSP for 3 cities (Nine neurons)

The fully connected Hopfield network is shown in figure-4 for 3 cities [11]. Here we use n^2 neurons in the network, where n is the total number of cities. The neurons here have a threshold and step-function. The inputs are given to the weighted input node. The network then calculates the output and then based on Energy function and weight update function, converges to the stable solution after some iteration. The most important task on hand is to find an appropriate connection weight. It should be such that invalid tours should be prevented and valid tours should be preferred.

The HNN can be best understood by the its energy function. Basically the energy function which was developed by Hopfield and Tank is used for the different problems. The energy function has various hollows that represent the patterns stored in the network. An unknown input pattern represents a particular point in the energy landscape and the pattern iterates its way to a solution, the point moves through the landscape towards one of the hollows. The iteration is carried on till some fixed number of time or till the stable state is reached. It can be given as below in the form of equation.

$$E = \underbrace{\frac{A}{2}\sum_{x}\sum_{i}\sum_{j\neq i} v_{x,i}v_{x,j}}_{E1} + \underbrace{\frac{B}{2}\sum_{i}\sum_{x}\sum_{y\neq x} v_{x,i}v_{y,i}}_{E2}$$
$$+ \underbrace{\frac{C}{2}\left(\sum_{x}\sum_{i} v_{x,i} - N\right)^2}_{E3} + \underbrace{\frac{D}{2}\sum_{x}\sum_{x\neq y}\sum_{i} d_{x,y}v_{x,i}\left(v_{y,i+1} + v_{y,i-1}\right)}_{E4}$$

Where
E1: Row inhibition, favor only 1 city in a row
E2: Column inhibition, favor only 1 city in a column
E3: Global inhibition, favor the state that all cities are present
E4: Distance inhibition, favor minimum distance of the tour
A,B,C,D,N: are constants

For the learning procedure the network must initialize the distance between cities and then repeat the iterations until the stopping condition is satisfied.

Intilization:
$u_{xi}(0) = u_{00} + ((rand-1)/10.*u_0)$

Stopping Condition:
$u_{xi}(t+1) = u_{xi}(t) + \Delta t(du_{xi}/dt)$

◆

$$\left[\frac{du_{xi}}{dt} = -u_{xi} - A\sum_{j\neq i} V_{xj} - B\sum_{y\neq x} V_{yi}\right]$$
$$- C(\sum_{x}\sum_{j} V_{xj} - N) - D\sum_{y} d_{xy}(V_{y,i+1}, V_{y,i-1})$$

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

V. Result & Discussion

As a result we have simulated the problem of six nodes as seen in figure-1. We applied the TSP using the Hop nets taking 36 neurons in all. We found that Hopfield is suitable for this kind of problem and hence we found the different tours. Tours are nothing but basically are the possible Hamilton circuit or shortest path. Different circuits are found in figure-5, figure-6, and figure-7 as per the following weights.

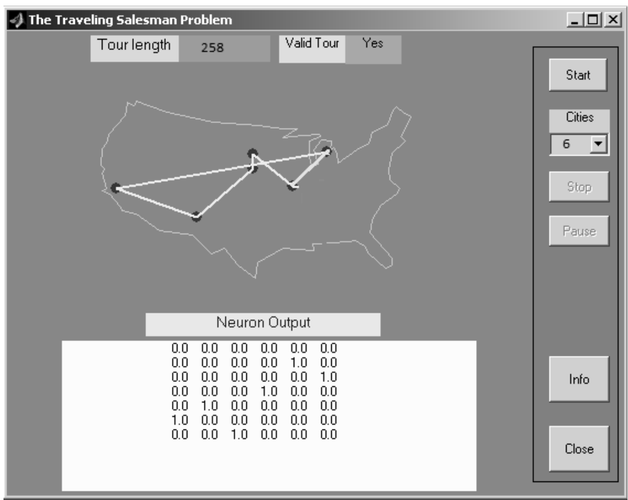


Figure 5. Path obtained by Matlab

For figure 5 the Hamilton circuit is HDTGPCH and Weight (HDTGPCH) = 20 + 45 + 71 + 22 + 58 + 54 = 270

For figure 6 the Hamilton circuit is HDTPCGH), Weight (HDTPCGH) = 20 + 45 + 67 + 58 + 36 + 32 = 258

For figure 7 the Hamilton circuit is Hamilton circuit- HCDTPGH), Weight (HCDTPGH) = 54 + 50 + 45 + 67 + 22 + 32 = 270

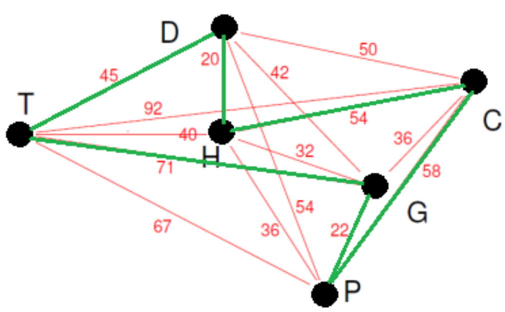


Figure 5. Path-1 (Hamilton circuit- HDTGPCH),
Weight (HDTGPCH) = 20 + 45 + 71 + 22 + 58 + 54 = 270

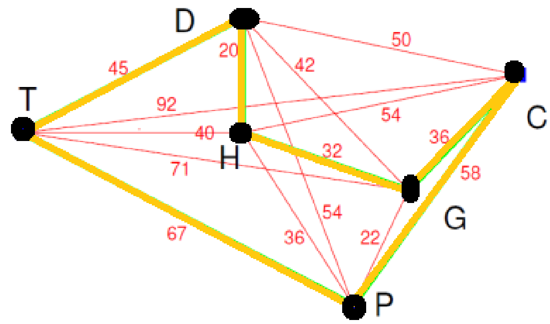


Figure 6. Path-2 (Hamilton circuit- HDTPCGH),
Weight (HDTPCGH) = 20 + 45 + 67 + 58 + 36 + 32 = 258

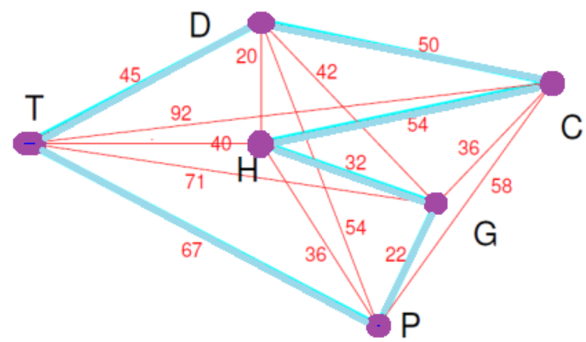


Figure 7: Path-1 (Hamilton circuit- HCDTPGH),
Weight (HCDTPGH) = 54 + 50 + 45 + 67 + 22 + 32 = 270

VI. Conclusion

In this paper the optimal route is successfully achieved using an efficient and effective route selection technique. And Hopfield neural networks for Travelling Salesman Problem has been successfully applied on six nodes which are located at six different positions. Earlier the similar problems were handles using ns-2, ns-3 etc. In future the same method can also be applied for more nodes but It will increase the number of neurons as well as the complexity. The network exhibits good performance in escaping from local minima of the energy surface of the problem. With a judicious choice of network internal parameters nearly 100% convergence to valid tours is achieved. Extensive simulations indicate that the quality of results is independent on the initial state of the network. The method proposed in this paper is a simple basic one. More enhancements can be achieved by developing the novel form of a utility function that better characterizes the level of the link's satisfaction in MANETs, which need further research.

References

[1] Gangwar, Sanjeev; Pal, S; Kumar, K, "Mobile Ad Hoc Networks: A Comparative Study of QoS Routing Protocols", International Journal of Computer Science Engineering and Technology (IJCSET), ISSN: 2231-0711, Vol:2, Issue:1, pp 771-775, 2012.

[2] CHAN-HO MIN and SEHUN KIM, "On-Demand Utility-Based Power Control Routing for Energy-Aware Optimization in Mobile Ad Hoc Networks", Wireless Personal Communications (2007) 41: 259–280, Springer, 2006.

[3] Ashish Kots, Manoj Kumar, "The fuzzy based QMPR selection for OLSR routing protocol", Published online: 27 April 2013, Springer Science+Business Media New York 2013.

[4] Zhihao Guo, Shaya Sheikh, Camelia Al-Najjar, Hyun Kim, Behnam Malakooti, "Mobile ad hoc network proactive routing with delay prediction using neural network" Published online: 27 October 2009_ Springer Science+Business Media, LLC 2009

[5] Abolhasan, M., Wysocki, T., & Dutkiewicz, E. (2004). A review of routing protocols for mobile ad hoc networks. Ad Hoc Networks, 2(1), 1–22.

[6] Seyed Saeed Sadat Noori, Seyed Ali Sadat Noori, Seyed Morteza Lari Baghal, "Optimization of Routes in Mobile Ad hoc Networks using Artificial Neural Networks", International Journal of Soft Computing And Software Engineering (JSCSE), ISSN: 2251-7545,Vol.2,No.4, 2012.

[7] Fatta, G., et al., 2003. A genetic algorithm for the design of a fuzzy controller for active queue management. IEEE transactions on systems, man, and cybernetics-part-C: applications and reviews, 33 (3), 313–324.

[8] Wu, Q., et al., 2004. On computing mobile agent routes for data fusion in distributed sensor networks. IEEE transactions on knowledge and data engineering, 16 (6), 740–753.

[9] Hopfield, J.J. and Tank, D.W., 1986. Neural computations of decisions on optimization problem. Biological cybernetics, 52, 141–152.

[10] Yip, P.P.C. and Pao, Y.-H., 1995. Combinatorial optimization with use of guided evolutionary simulated annealing. IEEE transactions on neural networks, 6 (2), 290–295.

[11] Ritesh Gandhi, "Approximate solution for the Traveling Salesman's Problem Using Continuous Hopfield Network", FINAL PROJECT REPORT (Fall 2001).

[12] Wen Liu, Lipo Wang, "Solving the Shortest Path Routing Problem Using Noisy Hopfield Neural Networks", *CMC*, 2009, Communications and Mobile Computing, International Conference on, Communications and Mobile Computing, International Conference on 2009, pp. 299-302, doi:10.1109/CMC.2009.366.

[13] C. W. Ahn, R. S. Ramakrishna, C. G. Kang, and I. C. Choi, "Shortest path routing algorithm using Hopfield neural network," *Electron. Lett.*, vol. 37, no. 19, pp. 1176–1178, Sep. 2001.

[14] A. W. Mohemmed, N. C. Sahoo, and T. K. Geok, "Solving shortest path problem using article swarm optimization," *Appl. Soft Comput.*, vol. 8, no. 4, pp. 1643–1653, Sep. 2008.

[15] Teng Gao ,Ren Cheng Jin ,Jin Yan Song ,Tai Bing Xu ,Li Ding Wang,‖ Energy-Efficient Cluster Head Selection Scheme -Based on Multiple Criteria Decision Making for Wireless Sensor Networks, Springer April 2012, Volume 63, Issue 4, pp 871-894

[16] C. Perkins and E. Royer, "Ad hoc on-demand distance vector routing, "in Proceeding of 2nd IEEE Workshop Mobile Comput. Syst. Appl. ,1999, pp. 90-100.

[17] H. XIAO, Winston K.G. SEAH, Anthony LO and Kee Chiang CHUA" Flexible Quality of Service Model for Mobile Ad-HOC Networks, " In proceedings of the IEEE Vehicular Technology Conference, May 2000.

[18] P. K. Suri, M.K. Soni and P. Tomar "Cluster Based QoS Routing Protocol for MANEI'" International Journal of Computer Theory and Engineering, Vol. 2 No. 5, October 2010.

[19] S. -J. Lee and M. Geria, "Dynamic Load-Aware Routing in Ad Hoc Networks," in Proceeding of IEEE ICC'OI, June 2001.

[20] H. Hassanein and A. Zhou, "Routing with Load Balancing in Wireless Ad Hoc Networks," in Proc. ACM MSWiM, 200l.

[21] K. Wu and J. Harms, "Load-Sensitive Routing for Mobile Ad Hoc Networks," in Proceeding of IEEE ICCCN'OI, Scottsdale, AZ, 2001.

[22] S-T. Sheu and 1. Chen, "A Novel Delay-Oriented Shortest Path Routing Protocol for Mobile Ad Hoc Networks," in Proceeding of IEEE ICC'OI, Helsinki, June 200l.

[23] E. Christopher Siddarth and K. Seetharaman, "A Cluster Based QoS Aware Service Discovery Architecture using Swarm Intelligence" in proceedings of IEEE International Conference on Computational Intelligence and Computing Research, 2012.

A Modified Black hole-Based Task Scheduling Technique for Cloud Computing Environment

Fatemeh Ebadifard
Department of computer
Iran University of science and technology
Tehran, Iran

Zeinab Borhanifard
Department of computer
Pooyesh Higher education institute
Qom, Iran

Ahmad Akbari
Department of computer
Iran University of science and technology
Tehran, Iran

***Abstract*— The issue of scheduling is one of the most important ones to be considered by providers of the cloud computing in the data center. Using a suitable solution lets the providers of cloud computing use the available resources more. Additionally, the satisfaction of clients is met through provision of service quality parameters. Most of the solutions for this problem aim at one of the service quality factors and in order to achieve this goal, variety of methods are used. Using the algorithm of modified black hole in this paper, a proper solution is presented to tackle the problem of scheduling the affairs in cloud environment. The proposed method reduces makespan, increases degree of load balancing, and improves the resource`s utilization by considering the capability of each virtual machine. We have compared the proposed algorithm with existing task scheduling algorithms. Simulation results indicate that the proposed algorithm makes a good improvement regarding the makespan and amount of resource utilization compared to schedulers based on Random assignment and particle swarm optimization Algorithms.**

***Keywords- cloud computing; task scheduling; Black hole; makespan; resource utilization*.**

I. INTRODUCTION

Cloud computing is one of the fields that has drawn attention of lots of the users in recent years. This is due to significant advantages that cloud services prepare for users in terms of cost and efficacy. The cloud environment provides a bed of servers in data center to provide the users sharing them as soon as they request for the resources. Service providers can provide the users with variety of services, renting virtual machines of cloud providers. Since providers of different services achieve the necessary virtual machines through cloud providers, the basic challenge of the service providers is presenting an effective method for scheduling the tasks in a way that they can provide the service quality necessary factors of aligned with needs of the service providers and users. Cloud task scheduling means optimized allocation of the requests to the computational resources available in data centers. When taking about scheduling, different kinds of virtual machines with specified constrains are presented for the users and service providers. Generally speaking, the scheduling algorithms are divided to two categories of static and dynamic ones:

Static scheduling: in algorithm of static scheduling, allocating tasks to the virtual machines takes place based on the capabilities of virtual machines and the primary status of each machine. In another words, this process is only based on primary information related to the nodes and their characteristics. This information include the amount of processing power, internal storage and the capabilities of storing and other the power of integration among other virtual machines as well. The important feature of the dynamic algorithms is that these algorithms do not regard all of the changes taking place dynamically in virtual machines. Moreover, they do not have are not adapted to the change of work load in virtual machines over time.

Dynamic scheduling: unlike static algorithms, in dynamic methods, in addition to primary capabilities of the each virtual machine, they assign the tasks to the virtual machines based on the existing status of that machine and work load assigned to it and according to the results of these evaluations, they transfer the requests from a machine to another one. Although these methods are more complicated than static method but they have more efficacies [1].

Using the black hole algorithm in this paper [2] improved by the Simulated Annealing, a method is

presented for scheduling the requests on virtual machines that in addition to the reducing the makespan time, increases the resource utilization. The advantage of the presented algorithm in above method comparing the previous ones is its simplicity and efficacy. The simulation results indicate that comparing the other heuristic algorithm such as PSO, above method could have better improvement in the makespan besides the efficacy of the resources. Briefly one can say that our main concentration in this paper includes the following:

- Presenting a suitable method for scheduling the requests using the modified black hole algorithm via simulated annealing, aiming at reduction of the makespan as well as Increase the utilization of the resources
- Using the fitness function to distribute the requests fairly
- The purposeful analysis and evaluation to show the amount of the efficacy of black hole algorithm in comparison to the other heuristic algorithms such as PSO

The rest of the paper consists of the following sections: in section 2 the related works are investigated; in section 3, after introducing the black hole algorithm, the mathematical model of algorithm is defined and then details of the suggested method are shown in details. In section 4 the simulation and evaluation of the suggested method are defined and finally in section 5, the conclusion and future works are presented.

II. Related Works

In cloud computing terminology, optimal assignment of requests to data center recourses is called task scheduling. Requests are assigned to different kinds of resources considering the service they might need. Task scheduling is one of the most important problems in the cloud computing. There are many studies regarding scheduling of tasks in cloud computing. In what follows, we will discuss some of these methods.

In 2010, Fang [5] proposed a scheduling method with the goal of increasing load balancing. In the proposed method, scheduling is performed on two level. On the first level, tasks are sent to appropriate virtual machines and if the machine is not efficient enough, it is placed on an appropriate physical machine. Simulation results indicate that this method provides good improvements in makespan and utilization of processor.

In 2010, Wang [3] proposed a method in which task scheduling is performed by a combination of Opportunistic Load Balancing (OLB) and Min-Min Load Balancing (MMLB). The combination of these two method reduces execution time and improves load balancing in the system. Moreover, the min-min scheduling algorithms minimizes the execution time of a task, which is possible through reducing the execution time of all tasks. The combination of these two algorithms maximizes efficient resource usage and increase the performance of task execution.

In 2012, Krishna et al. [6] proposed a scheduling method based on load balancing with the goal of reducing waiting time and increasing response time. The proposed method was inspired by the honey bee behavior. After assigning tasks, the proposed method divides virtual machines into three groups of under loaded, balance, and over loaded and if a machine is over loaded, tasks are transferred from that machine to under loaded one. Simulation results indicate that this method makes good improvements in makespan and response time. Moreover, it increases the degree of load balancing.

In 2012,Zhan et al. [12] proposed a mixed method using both PSO and Simulated Annealing (SA) algorithms. The proposed method tends to reduce tasks' average execution time and increase convergence speed to the optimal solution. This study shows that the improved PSO-based method has better efficiency comparing to genetic algorithm, simulated annealing, and ant colony; although, combining PSO and SA algorithms results in more computational complexity. Mixing PSO with SA algorithms, Kaur and Sharma [8] proposed a new task scheduling method for cloud environment. Their primary goal was to optimize resource utilization and maximize providers' profit.

In 2014,Abdi et al. [11] proposed an improved PSO-based task scheduling method using "shortest job to fastest processor" algorithm for generating the initial population. The method's goal was to reduce makespan time. Their results show that the proposed method has lower makespan time comparing to GA-based and simple PSO-based solutions.

In this paper, we propose a Black hole-based task scheduling algorithm by benefiting from SA algorithm to generate a more suitable initial population and choosing a more appropriate goal function. Our study shows that Black hole method has not been used to solve the problem of task scheduling in the cloud computing. Despite all the above mentioned works which consider either clients' benefits or providers' benefits, our method decreases makespan duration and increases resource utilization at the same time, and it can meet providers' and clients' needs simultaneously.

III. Modified Black hole Based Task Scheduling

The Black hole-based task scheduling method is explained in this section. To bring a Mutual benefit to both providers and clients, the proposed method aims to maximize resource utilization and minimize the makespan time. First, we have briefly described Black hole technique and its components; then, we have presented the problem formulation and our proposed method in details.

3.1 Classic Black hole

Black hole is a new meta-heuristic algorithm introduced by Hatamlou [2] at 2012. It is a population-based method that has some common features with other population-based methods. The BH algorithm is more similar to the natural black hole phenomenon and evolving of the population is done by moving all the candidates towards the best candidate in each iteration, namely, the black hole.

Like other population-based algorithms, in the proposed black hole algorithm (BH) a randomly generated population of stars can be considered as a possible solution for the problem, and it searches the problem space for the optimized solution. Using a fitness function, performance of each star is evaluated at the end of every iteration according to Eq. (3-6) and the best candidate in the population, which has the best fitness value, is selected to be the black hole and the rest form the normal stars. After initializing the black hole and stars, all the stars start moving towards the black hole according to Eq. (3-1).

(3-1)

$$x_i(t+1) = x_i(t) + rand \times (x_{BH} -, x_i(t)) \quad i = 1, 2, \ldots, N$$

Where xi(t) and xi(t + 1) are the locations of the ith star at iterations t and t + 1, respectively. xBH is the location of the black Hole in the search space. Rand is a random number in the interval [0, 1]. N is the number of stars.

In Every iteration if a star reach a location with lower fitness than the black hole, this star replace with black hole and then stars start moving towards this new location. Every star (candidate solution) that crosses the event horizon of the black hole will be sucked by the black hole. Every time a candidate (star) dies another candidate solution (star) is born and distributed randomly in the search space. The radius of the event horizon in the black hole algorithm is calculated using the following equation:

(3-2)

$$R = \frac{f_{BH}}{\sum_{i=1}^{N} f_i}$$

Where fBH is the fitness value of the black hole and fi is the fitness value of the ith star. N is the number of stars (candidate solutions).

When the distance between a candidate solution and the black hole (best candidate) is less than R, that candidate is collapsed and a new candidate is created and distributed randomly in the search space. The pseudo code of Modified black hole task scheduling is presented in Figure 1.

```
Initialize a population of stars with random locations in the search space

Loop

For each star, evaluate the objective function

Select the best star that has the best fitness value as the black hole

Change the location of each star according to Eq. (3-1)
```

Figure 1 the black hole algorithm pseudo code [2]

3.2 Task-Resource Scheduling Formulation

Different objectives can be considered for scheduling tasks in cloud computing environment. Minimizing the makespan time and maximizing resource utilization is our focus of attention in our method.

To formulate the problem we have denoted a set of tasks $Task = \{T_1, T_2, T_3, \ldots, T_i\}$ where $i \in \{1, 2, \ldots, n\}$. Tasks are assumed to be non-preemptive and independent. We have defined a set of m virtual machines, $VM = \{VM_1, VM_2, VM_3, \ldots, VM_j\}$ interconnected by network where $j \in \{1, 2, \ldots, m\}$. The tasks will be processed on virtual machines. Completion time and processing time of task T_i on virtual machine VM_j are denoted as CT_{ij} and PT_{ij} respectively. The objectives are minimizing the overall task completion time and maximizing the average resource utilization. overall task completion time is called makespan and is defined by Eq. (3-3) which is extracted from [6]:

$$Makespan = \max\{CT_{ij} \mid i \in T, i = 1, 2, \ldots, n \text{ and } j \in VM, j = 1, 2, \ldots, m\} \quad (3-3)$$

Virtual machines have its own processing unit, and processing time of each specific task on each specific VM is supposed to be known; therefore, utilization of each resource is defined by Eq. (3-4), and average utilization is defined by Eq. (3-5):

$$Utilization_{VM_j} = \frac{\sum_{i=1}^{n} PT_{ij}}{makespan} \quad (3-4)$$

$$Average\ Utilization = \frac{\left(\sum_{j=1}^{m} Utilization_{VM_j}\right)}{m} \quad (3-5)$$

Regarding our objectives, the fitness function is defined by Eq. (3-6):

$$Fitness\ Function = \frac{(\min)\ makespan}{(\max)\ Average\ utilization} \quad (3-6)$$

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Eq. (3-6) shows that a star has a better position if it has a lower fitness value, and this star can increase average resource utilization and decrease makespan time. The details of the proposed method is explained in the next subsection.

3.3 Proposed Black hole-Based Task Scheduling Method

Detailed explanation of the proposed method and pseudo code of its algorithm are presented in this section. The steps of the algorithms are as following:

Step 1: Defining of Stars and position vectors for a problem with n tasks, we have defined each star as a N-dimensional vector $X = (X_1, X_2, \dots, X_n)$, where $X_i \left(i \in \{1, 2, \dots, n\}\right)$ represents index of a virtual machine on which task i will be processed. Position of stars are initialized randomly. For example for a 6 task and 3 virtual machine problem a star's position vector can be initialized as shown in Table. 1.

Table 1 Values of a randomly initialized star

Task 1	Task 2	Task 3	Task 4	Task 5	Task 6
VM 2	VM 1	VM 1	VM 3	VM 2	VM 2

Step 2: In standard Black hole algorithm initial stars are created randomly, but randomness decreases the chance of algorithm to converge to best solution, in order to improve the behavior of Black hole algorithm, we merge Simulated Annealing algorithm(SA) into Black hole, i.e. instead of generating initial population randomly we Improve them considering SA algorithm. All other steps are similar to standard Black hole algorithm.

Step 3: Calculating fitness function and Specifying black hole

Fitness value of each Star will be calculated using Eq. (3-6). Comparing the current fitness value of the whole population together, the lowest fitness value will be specified as black hole best position.

Step 4: Updating stars' position

The position vector of each star will be updated respectively. All the stars start moving towards the black hole according to Eq. (3-1).

Step 5: Terminating condition

Steps 3 and 4 will be repeated until the maximum number of iterations is reached.

Pseudo code of our proposed algorithm is shown in Fig. 2. Position initialization of our method has an advantage over the base Black hole task scheduling method because base Black hole task scheduling method initializes stars' position vectors randomly, but our proposed method executes a load balancing step after random initialization to improve the stars' positions. Therefore, each initial star can present a proper solution which will be improved by moving in the problem space. The simulation results, presented in section 4, show that comparing to the base black hole and the PSO-based task scheduling method, our method leads to more efficient completion time and resource utilization.

```
Input:

Task = {T1, T2, T3, ..., Ti}
VM = {VM1, VM2, VM3, ..., VMj},

    Output: best position of Tasks on the VM

Start:

 1: Set star dimension equal to the size of ready tasks, Initialize
 stars position randomly.

 2 : for each star run Simulated Annealing algorithm
for balancing star position using fig 3

 3: For all Stars, calculate its fitness value by in Eq. 3-6

       If ( fitness value < bh-fitness)

          set the current fitness value as the new bh-fitness

 4: For all stars, update their positions using Eq. (3-1)

 5: For all Stars, calculate R= distance between star positon
and black hole position and R by using Eq. (3-2)

       If (result<R)

        Remove current star and replace it with a new star in a
      random location in the search space

 6: repeat steps 3 to 6 If termination criteria or maximum
iteration is not satisfied.
```

Figure 2 the Modified black hole-based task scheduling pseudo code

```
Simulated Annealing
 S = Choose an initial solution
 T = Choose an initial temperature
 REPEAT
   S' = Generate a neighbor of the solution S
   ΔE = objective(S') – objective(S)
   IF (ΔE > 0) THEN // S' better than S
     S= S'
   ELSE with probability EXP (ΔE/ T)
     S = S'
   END IF
   T = lower the T using linear/ non-linear techniques
 UNTIL meet the stop criteria
End
```

Figure 3 The Simulated Annealing pseudo code

IV. Simulation Results

In this section, we have presented the simulation setup and evaluation of the proposed algorithm based on the results of conducted tests. It is worth to mention that the experiments are conducted in SaaS level. We have used CloudSim [14] to simulate cloud computing SaaS level. This simulator gives capability of simulating a virtualized environment, and it supports on demand resource provisioning. We have extended the CloudSim simulator for modeling our method. The purpose of our method, as mentioned before, is to present an efficient Black hole-based task scheduling algorithm for cloud computing environment which can reduce completion time of the longest task (makespan) and increase the average resource utilization of the cloud data center. We have defined 10 stars for the Black hole population. The termination condition is set to 100 iterations. To analyze our method, several experiments are conducted in two different test setups. Test setups and experiments are described in the rest of this section. The input parameters and variables used in the task scheduling problem are presented in Table 2.

Table 2 The input parameters and variables for problem

Particle Swarm Optimization algorithm(PSO)	initial population	10
	Iteration Number	100
	C_1 و C_1	1.49445
	r_1 و r_2	Random number between 0 and 1
Black hole algorithm	initial population	5
	Iteration Number	100
	rand	Random number between 0 and 1
Simulated Annealing algorithm(SA)	Iteration Number	100
	Initial Temp	1000

- **Test setup 1**

In the first test setup, we have defined a cloud data center which has three hosts, each capable of supporting virtualization technology and sharing its resources among several virtual machines. Hosts' hardware specifications are presented in table. 3. Sixteen virtual machines are supposed to be running on these three hosts; each of them has a distinct specifications. Each virtual machine executes applications with different number of instructions varying from 500 to 4500. We have used simulated workload for these series of tests.

Table 3 Hosts' technical specifications

HostId	Number of processing Cores	Processing speed (Mips)	Ram (MB)	Hard (MB)	Bandwidth (Mbps)
1	4	5000	204800	1048576	102400
2	2	25000	102400	1048576	102400
3	1	10000	51200	1048576	102400

The results of the proposed method are compared to four other methods: 1) simple Round Robin (RR) algorithm, 2) classic PSO-based task scheduling method which initializes particles' position and velocity vectors randomly, 3) classic Black hole method, and 4) Modified Black hole -based method which aims to improve Black hole algorithm convergence speed taking advantageous from SA algorithm.

Fig 4. Compares makespan of these four methods for different number of tasks. The result illustrates that our proposed method outperforms other four methods. Our method is qualified to balance the load of Stars in the beginning of first iteration, therefore it provides a better makespan comparing to the other methods specially with more tasks. The proposed method affects the makespan time more effectively than others.

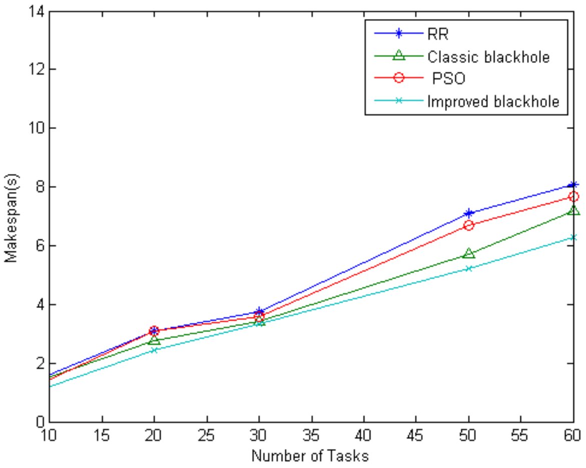


Figure 4 makespan comparison

Fig. 5 represents response time for different number of task for the Round Robin, PSO, classic Black hole, and the proposed algorithms. It is clear that our method provides a better response time in comparison to three other methods.

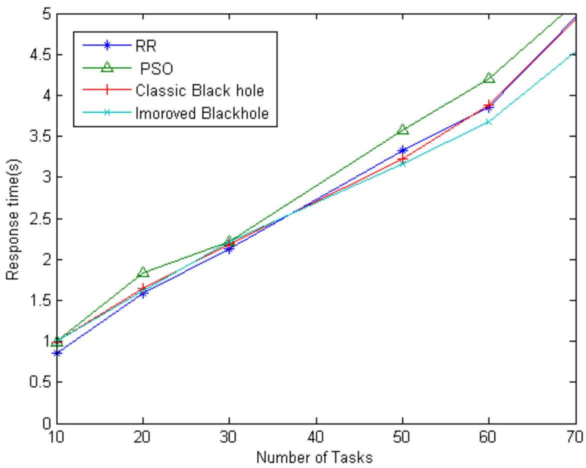


Figure 5 response time comparison

Comparison of resource utilization among the proposed method, PSO algorithm and RR algorithm is shown in Fig. 6. It is evident that both Modified Black hole-based and PSO-based task scheduling methods results in higher resource utilization comparing to the RR algorithm. Because of its ability of balancing the load of virtual machines, our proposed method has better performance in comparison with the classic PSO algorithm.

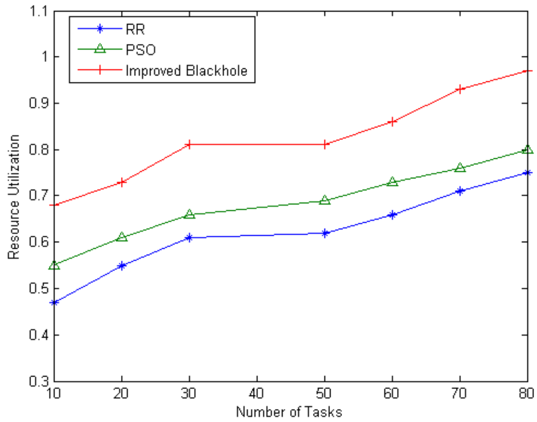


Figure 6 Resource utilization comparison

- **Test setup 2**

To Analyze performance of the proposed method under real workloads, we have conducted another experiment with a new test configuration. Four hosts, with technical specifications provided in table. 4, are supposed to work in a cloud data center. 40 homogeneous virtual machines are supposed to be running on these hosts. Hardware specification of virtual machines are presented in table. 5.

Table 4 Hosts' technical specifications

CPU	Numb er of cores	Processin g speed (Mips)	Ram (MB)	Hard (MB)	Bandwi dth (Mbps)
Core_2_E xtreme_X6 800	2	27079	2048 0	104857 6	102400
Core_i7_E xtreme_Ed ition_3960 X	6	177730	1024	104857 6	102400
Core_i7_E xtreme_Ed ition_980 X	6	147600	2048 0	104857 6	102400
Core_i7_8 75K	4	92100	2048 0	104857 6	102400

Table 5 Virtual machines' technical specification

CPU	Number of cores	Processing speed (Mips)	Ram (MB)	Hard (MB)	Bandwidth (Mbps)
Core_i4_Extreme _Edition	1	9726	512	10240	1024

We have used a workload which is logged by NASA Ames Research Center from October to December in 1993 [15]. It contains 42240 jobs. Each job is converted to a Cloudlet regarding to its completion time and processing rate of existing processors. Each Cloudlet is a task that can be used in CloudSIM simulator.

Fig. 7 shows makespan comparison of four algorithms. With fixed number of virtual machines (40 virtual machines), number of jobs is increased from 300 to 2500. It is evident that our proposed method works efficiently under real workloads and outperforms other three algorithms in terms of makespan duration.

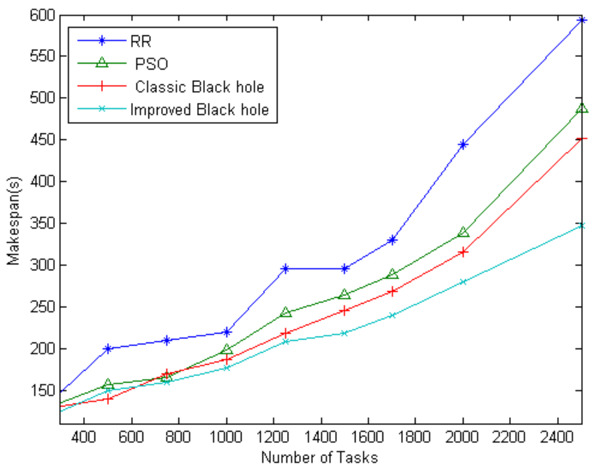


Figure 7 Makespan time comparison with real workload

In addition, The Simulation results illustrate that our proposed method increases resource utilization in the entire system, as well as decrease makespan, and increase response time. Moreover, the proposed method also is more efficient from computational complexity point of view and has simple implementation in comparison with other heuristic algorithms.

V. Conclusion and Future Work

In this paper, a method of scheduling based on black hole algorithm was presented. Using this algorithm and choosing a suitable fitness function, increment of the resources utilization and the reduction of the makespan are feasible. Then proposed method with three algorithms of round robin algorithm as well as algorithms of PSO [13] and base black hole as one of the heuristic algorithms suitable for the issue of scheduling were compared. The simulation results show that in spite of the computational simplicity, the modified black hole algorithm has a good optimization in terms of makespan and resource utilization. Our method is generic and scalable as it can be deployed in data centers with any number of tasks and resources by increasing task-resource array dimension. Furthermore, our method is applicable for any cloud environments with independent and non-preemptive tasks. In the future we plan to expand our method for workflow applications and taking other QoS criteria like fault tolerance capability and cost reduction into account.

References

[1] K. A. Nuaimi, N. Mohamed, M. A. Nuaimi, and J. Al-Jaroodi, "A survey of load balancing in cloud computing: Challenges and algorithms," in Network Cloud Computing and Applications (NCCA), 2012 Second Symposium on, 2012, pp. 137-142.

[2] Abdolreza Hatamlou, "Black hole: A new heuristic optimization approach for data clustering", Information Sciences , journal homepage: www.elsevier.com/locate/ins ,2013. pp.175–184.

[3] L. Wang and L. Ai, "Task Scheduling Policy Based on Ant Colony Optimization in Cloud Computing Environment," in LISS 2012, ed: Springer, 2013, pp. 953-957.

[4] S. K. Chaharsooghi and A. H. M .Kermani, "An effective ant colony optimization algorithm (ACO) for multi-objective resource allocation problem (MORAP)," Applied Mathematics and Computation, vol. 200, pp. 167-177, 2008.

[5] Y. Fang, F. Wang, and J. Ge, "A task scheduling algorithm based on load balancing in cloud computing," in Web Information Systems and Mining, ed: Springer, 2010, pp. 271-277.

[6] P. V. Krishna, "Honey bee behavior inspired load balancing of tasks in cloud computing environments," Applied Soft Computing, vol. 13, pp.2013 ,2303-2292 .

[7] X. Nan, Y. He, and L. Guan, "Optimal resource allocation for multimedia application providers in multi-site cloud," in Circuits and Systems (ISCAS), 2013 IEEE International Symposium on, 2013, pp. 449-452.

[8] G. Kaur and E. S. Sharma, "Optimized Utilization of Resources Using Improved Particle Swarm Optimization Based Task Scheduling Algorithms in Cloud Computing," International Journal of Emerging Technology and Advanced Engineering, vol. 4, June 2014 2014.

[9] S. Pandey, L. Wu, S. M. Guru, and R. Buyya, "A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments," in Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on, 2010, pp. 400-407.

[10] T. Kokilavani and D. D. G. Amalarethinam, "Load balanced min-min algorithm for static meta-task scheduling in grid computing," International Journal of Computer Applications, vol. 20, pp. 43-49, 2011.

[11] S. A. M. Solmaz Abdi, Saeed Sharifian, "Task Scheduling using Modified PSO Algorithm in Cloud Computing Environment," presented at the International Conference on Machine Learning, Electrical and Mechanical Engineering (ICMLEME'2014), Dubai (UAE), 2014.

[12] S. Zhan and H. Huo, "Improved PSO-based task scheduling algorithm in cloud computing," Journal of Information & Computational Science, vol. 9, pp. 3821-3829, 2012.

[13] k. krishnasamy, "task scheduling algorithm based on hybrid particle swarm optimization in cloud computing environment," Journal of Theoretical & Applied Information Technology, vol. 54, 2013.
[14] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. De Rose, and R. Buyya, "CloudSim: a toolkit for the modeling and simulation of cloud resource management and application provisioning techniques," Journal of Software: Practice and Experience, 2010.
[15] D. G. Feitelson and B. Nitzberg, "Job characteristics of a production parallel scientific workload on the NASA Ames iPSC/860," in Job Scheduling Strategies for Parallel Processing, 1995, pp. 337-360

A multicast routing protocol based on ODMRP with Stable link in mobile ad hoc networks

Ebrahim Asadi
Department of Computer Engineering
Shabestar Branch,Islamic Azad University
Shabestar, Iran

Ali Ghaffari
Department of Computer Engineering,
Tabriz Branch,Islamic Azad University
Tabriz, Iran

***Abstract*—Mobile ad hoc networks are more flexible than tradition networks since they do not require fixed infrastructure and allow all nodes move in a random trajectory, which leads frequent rerouting and degrades network performance. So, an important issue in mobile computer network research is routing in mobile ad hoc networks. Multicast sending is one of the methods used for routing in mobile ad hoc networks because of its group activities. However, some problems exist in multicast sending. For example, when receiver nodes attempt to send acknowledgments or path repetition packets simultaneously, crashes may occur, which leads to packet loss. On the other hand, link expiration is another reason for packet loss. In this study, a multicast routing protocol is offered, which uses a combination of two parameters of the received signal's power and the remaining energy to estimate the stability of the link. SINR is used at each node in conjunction with various transmitters to determine a reliable path that reduces link failure and end-to-end delay. The aim is to find the best link with probability of the highest life cycle for each path. Simulation results of the proposed method using NS-2 simulator indicate the good performance of IMP-ODMRP measures in packet delivery rate, end-to-end delay, packet loss rate, and packet collision rate.**

Keywords-Mobile ad hoc networks; multicast; routing; IMP-ODMRP protocol; Standard ODMRP; Stable Link.

I. INTRODUCTION

All MANET is a mobile ad hoc, temporary, and instantaneous network that is developed for special purposes. Indeed, wireless networks are a collection of wireless mobile nodes that are infrastructureless, autonomous, and without any centralized management network. Therefore, nodes in this type of network are responsible for dynamically discovering each other. Based on the nature of dynamics, the network topology of this type of network changes continuously. Because MANETs are mobile, connection changes are unpredictable. The biggest challenge encountered by these kinds of networks is maintaining routing packet efficiency until reaching the destination without creating overhead [18]. Consequently, some methods must be proposed to incorporate less overhead. Several routing algorithms are presented by MANET networks, each of them having different features, advantages, and disadvantages. There are various methods of classifying routing protocols in mobile ad-hoc networks; however, most of them depend on routing strategy and network structure [4], [10], [12]. In recent years, a number of multicast protocols for ad hoc networks have been proposed. Based on the routing structure, they can broadly be classified into two categories: tree-based and mesh-based. Tree-based MANETs require fundamental changes to conventional routing protocols for both unicast and multicast communication owing to their unique features. With the rapid growth of group communication services, the multicast routing in MANETs has attracted much attention recently [5]. In multicast routing, a path is set up connecting all group members so that bandwidth is not wasted. Group communication applications include audio/video conferencing as well as one-to-many data dissemination in critical situations such as disaster recovery or battlefield scenarios [22-25]. In addition, their applications are seen in mobile/wireless environments where mobility and topology changes produce very high overhead and affect throughput performance in terms of packet delivery ratio. Since group-oriented communication is one of the key application classes in MANET environments, a number of MANET multicast routing protocols have been proposed. These protocols are classified according to two different criteria. The first criterion maintains routing state and classifies routing mechanisms into two types: proactive and reactive. However, redundant routes cause low multicast efficiency. In this paper, we focus on the ODMRP protocol; ODMRP is a state-of-the-art on-demand multicast routing protocol [7], [8]. It is a mesh-based source-initiated protocol that uses the forwarding group concept to establish a mesh. Moreover, it follows the “soft state” approach to maintain a mesh. To overcome these limitations, the On-Demand Multicast Routing Protocol (ODMRP) was developed [14]. ODMRP is a mesh-based, instead of a tree-based, multicast protocol that provides richer connectivity among multicast members. By building a mesh and supplying multiple routes, multicast packets can be delivered to destinations in the face of node movements and topology changes. In addition, the drawbacks of multicast trees in mobile wireless networks (e.g. intermittent connectivity, traffic concentration, frequent tree reconfiguration, and non-shortest path in a shared tree etc.) are avoided. To establish a mesh for each multicast group, ODMRP uses the concept of forwarding group [16]. The forwarding group is a set of nodes responsible for forwarding multicast data on shortest paths between any member pairs. ODMRP also applies on-demand routing techniques to avoid channel overhead and improve scalability. A soft state Approach is taken to maintain multicast group members. No explicit control message is

required to leave the group. We believe the reduction of channel/storage overhead and the richer connectivity make ODMRP more attractive in mobile wireless networks. The remainder of this paper is organized as follows. In Section 2, related work is presented. Section 3, introduces general description multicast routing in mobile ad hoc networks, and Section 4 provides an explanation for on-demand multi cast routing protocol (ODMRP). Section 5 presents the proposed method, and Section 6 introduces the steps of the proposed method. Finally, Section 7 concludes the paper.

II. Related work

Lately, many researches have been carried out on MANETS based on multicasting. Multicasting is one of methods that because of group oriented computing are mostly used in MANET routing [14], [26-27]. Nowadays link stability and link quality is the major topic to be resolved and for this, the routing protocol is to prefer stable links than transient links. For link stability and link quality, signal strength is also measured. It is based on Route life time Assessment based routing [6] in which a link is considered to be stable if it exists for a time of at least Athresh= 2rtx/v, where rtx is the transmission range and v denotes the relative speed of two devices [1]. Signal Stability Adaptive Routing (SSA) [2] follows a similar approach. It distinguishes strongly connected from weakly connected links, where a link is considered to be strongly connected if it has been active for a certain predefined amount of time. The protocol, termed ODMRP (On-Demand Multicast Routing Protocol) is a mesh-based, rather than a conventional tree based multicast scheme and uses a forwarding group concept (only a subset of nodes forwards the multicast packets via scoped flooding). It applies on-demand procedures to dynamically build routes and maintain multicast group membership. ODMRP is well suited for ad hoc wireless networks with mobile hosts where bandwidth is limited, topology changes frequently, and power is constrained [11]. There exist many surveys of multicast routing protocols for Mobile ad hoc networks but only few of them have focused on QoS based multicast routing for MANETs. In [9], the authors attempted to describe QoS provisioning at the MAC layer. In [20], the authors proposed a QoS-based multicast routing protocol QMMRP. They used entropy of node and bandwidth reservation policy to find a stable link with sufficient bandwidth. In [3], the authors conducted an extensive survey on multicast routing protocols and focused on different QoS requirements by different applications of MANET. In [15], the authors used a stability function as the main path selection criterion based on the calculation of the mobility degree of a node relative to its neighbor. Their routing mechanism was based on link stability, which minimizes frequent path disconnections and guaranties other QoS requirements such as packet delivery ratio. In [21], the authors proposed a link stability estimation model based on received signal strength indication. They integrated this model into MAODV and presented a Stability-based Multicast Routing protocol termed SMR. SMR can discover more available and stable routes and better adapt to network topology changes. SMR is an extension of the MAODV protocol, thus the Multicast Routing Table, Route Request (RREQ), Route Reply (RREP) and Route Error packet formats are similar to those used in MAODV. Simulation results show the superiority of SMR over existing methods in terms of packet delivery ratio, average end-to-end delay and routing packet overhead. In [17], the authors propose an agent based Multi-Constrained QoS aware multicast routing scheme based on MAODV (MC_MAODV), which uses a set of static and mobile agents. It depicts the QoS multicast model with multiple constraints that may deal with bandwidth reservation, delay constraint and packet loss to multicast session. In [13], the authors proposed an approach for using multiple metrics simultaneously, with one of the metrics, which they call optimizable, reflecting consuming network resources, and other metrics, which they call restrictive, reflecting QoS requirements. If a route length goes beyond a threshold in at least one of the restrictive metrics, the route shall not be chosen for packet delivery to escape network resources waste. Thus, the best route is chosen in an optimizable metric in the class of routes allowed by restrictive metrics. The approach is applicable for both unicast and multicast traffic in MANETs. On the values from this watchdog, trust value on the neighbor is being increased or decreased dynamically. The method is implemented only on the ODMRP protocol.

III. General Description Multicast Routing Mobil Ad hoc Networks

The mobile ad hoc network is a format of mobile platforms in which a router with multiple hosts and wireless communication devices - herein simply presented as nodes - are free to move about arbitrarily. The nodes may be located in or on airplanes, ships, cars, perhaps even on people or very small devices, and there may be multiple hosts per router. A MANET is an autonomous system of mobile nodes. The system may operate in isolation, or may have gateways to and interface with a fixed network. In the latter operational mode, it is typically envisioned to operate as a "stub" network connecting to a fixed internetwork. Stub networks carry traffic originating at and/or destined for internal nodes, but do not permit exogenous traffic to "transit" through the stub network. MANET nodes are equipped with wireless transmitters and receivers using antennas, which may be omni-directional (broadcast), highly-directional (point-to-point) possibly steerable, or some combination thereof. At a given point in time, depending on the nodes' positions and their transmitter and receiver coverage patterns, transmission power levels and co-channel interference levels, a wireless connectivity in the form of a random, multihop graph or "ad hoc" network exists between the nodes. This ad hoc topology may change with time as the nodes move or adjust their transmission and reception parameters. There are many multicast protocols including MAODV, PIM, ADMR, PUMA, OBAMP, and ODMRP. Our research has been conducted using the latter. ODMRP is a mesh based routing protocol and uses the concept of forwarding group in which only a subset of nodes will be allowed to forward multicast packets. In this protocol, multiple routes are established and updated by source on demand where a source node broadcasts a JOIN-QUERY if it does not possess any routes to send its data packet. This JOIN-QUERY is periodically broadcasted to refresh the membership information and update routes. When an intermediate node receives a join-query, it stores the source ID and sequence

number in its message cache to detect any duplicates. When the join-query packet reaches a multicast receiver, it creates and broadcasts a JOIN-REPLY to its neighbors. When a node receives a join-reply, it checks whether the next hop node ID of one of the entries matches its own id or not. If it does, the node realizes that it is on the path to the source, and thus, it is part of a forwarding group [13], [21].

IV. EXPLANATION OF ON-DEMAND MULTICAST ROUTING PROTOCOL (ODMRP)

ODMRP is just a state-of-the-art on-demand multicast routing protocol. It is just a mesh based source initiated protocol. It uses the forwarding group concept to set up a mesh. It follows the “soft state” approach to keep a mesh[19].

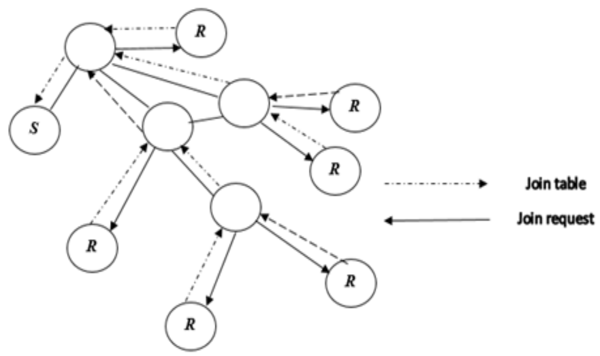
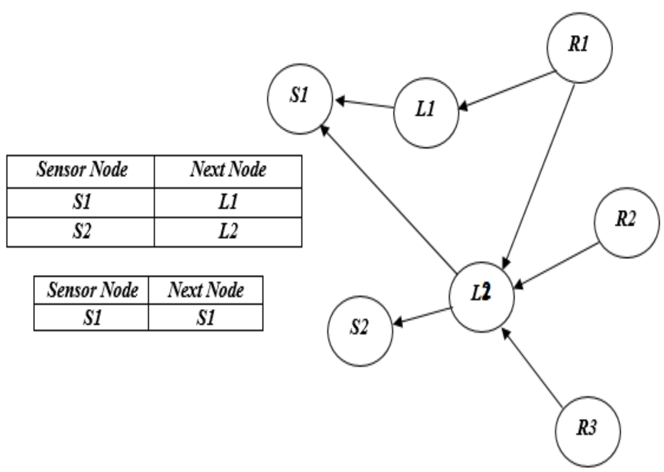


Fig. 1. Multicast Routing Protocol (ODMRP)

Figure 1 illustrates the on-demand procedure for membership setup and maintenance of the ODMRP. Whenever an Each time a source node desires to send data packets to the multicast group, it periodically broadcasts the JOIN_QUERY packet to the network and when received by each intermediate node, it checks whether the if the received packet is a duplicate or not based on the sequence number in the packet header header [19].



Sensor Node	Next Node
S1	L1
S2	L2

Sensor Node	Next Node
S1	S1

Fig. 2. Join table forwarding(ODMRP)

In Figure 2 Each time a source node desires to join or leave the group, it doesn't require any control packets. If aIn case a source node does have no data packet to send, it just stops sending any packets to the multicast group. There are You can find three kinds of tables in the ODMRP architecture, which are Member Node Table, Routing Table, and Forwarding Group Table. The Member Node table is used for storing the source information. Each entry in the table is designated by source ID and time of last JOIN_QUERY received pair. If JOIN_QUERY isn't received by way of a member node inside a refresh period, that entry is removed. The Routing Table is created is established on-demand and is maintained by each node [19].

V. DETAILS OF THE PROPOSED PROTOCOL IMP-ODMRP (OVER VIEW)

A. *Sustainability link*

The swing of link stability due to mobility or transmission media characteristics in wireless network influences the network's performance. The productivity of a dynamic routing protocol can be divided by ability to face link unreliability and routing overhead in terms of calculations and reconfiguration/ rerouting. Link stability as the base of every routing decision can lead to a protocol that includes the following capabilities: Effective energy: low overhead of communication and calculations that results from the definitive reduction in links by reduction of rerouting. Flexibility of movement: selected links in the case of long time communication disconnection are more resistant to node movements. Stability: the same paths are maintained for a long time to reduce the overhead on routing tables.

B. *Definition of the proposed parameters*

Noise rate (ζ): Noise rate is defined as the ratio of signal power (S) to a combination of noise power (N) and intervention (I) and can be presented mathematically as SINR = S/N + I. As the accurate sum of intervention and noise power cannot be calculated accurately, SINR is estimated using the average of receive during the rest period. SINR is used for defining the quality of network links and connections. The parameter of Remaining Energy Strength (RES): We calculate the needed energy for data transfer before sending based on the file or data size. Each node can be informed of the remaining energy strength and its status by GPS. In addition, beginning and source nodes for each packet can be calculated by Eq 1.

$$ETX= \frac{\text{Packet Size} * \text{PTX}}{\text{BW}} \qquad (1)$$

Note that Ptx is the needed power for the transfer of a packet and BW is the bandwidth of the link. The shortest path

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

from source node to target node needs a minimum energy. Finally, the minimum energy each node needs for data transfer can be calculated from Eq 2.

$$RES = n * (Etx + Erx) \qquad (2)$$

Note that n stands for the total amount of data packets for transmission, and Erx is the needed energy for receiving data packets.

C. Stable routes

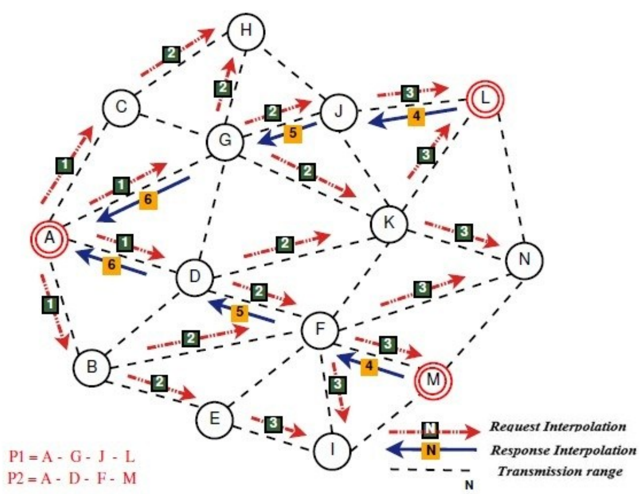


Fig 3. Work stream flow of standard ODMRP protocol

Link stability in the path is estimated from source to target for each link. To select the next node when creating a path, the link's stability is estimated by remaining energy and SINR. To ensure that gathering data of all existing links has been possible before selecting the best link, in node (A), each packet becomes buffered for expiration date. All the received data packets that are neighbors of A are examined during this time by SINR, and then, A selects the best link with the highest remaining energy and SINR rate. The proposed method selects the link with the highest SINR and highest remaining energy so that it has the highest possibility for keeping route for the longest time. The created path with packet storage in each receiver node is shown in Figure 3 Represents thickness of SINR arrow.

VI. THE PROPOSED METHOD

The various steps of the proposed method are as follows:

1- *The source of the packet creates a "join request" and sends it to the address of the multicast group.*
2- *Middle nodes receive the "join request" from other nodes.*
3- *Repeated packets of same source to the similar multicast address and same sequence number are discarded.*
4- *Receiver node calculates the remaining energy and SINR rate for various transmitters and saves node data in the related buffer.*
5- *Each node adds a timer, which was created at the start of receiving the first data packet. This packet is set to "expiration time" and only received packets are considered during this time.*
6- *After expiration time, each node calculates the remaining energy and the value of SINR for all links and selects one with the highest remaining energy. That node is set as the address of the next hop with the highest remaining energy and SINR rate.*
7- *The middle node then sends the packet to the target in a series of multi-cast packets.*

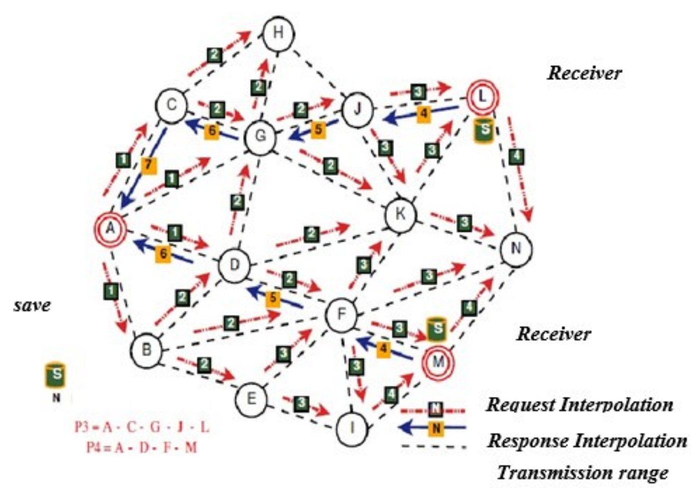


Fig 4. Work stream of the proposed method

According to the ODMRP protocol, the least number of step counting is considered for defining the path between source and target. In this method, we used the two parameters of remaining energy and signal power for defining the stable

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

path. Figure 3 shows the paths P1 and P2 between source A and target L and M, which is achieved using minimum step number while P3 and P4 is shown in Figure 4 using the sum of remaining energy of path nodes and node signal power. Path P3 has changed from path P1 according to the lower signal power from node A. We did not see any change in the path of P4 to P2 because it is on the exposure of best-received signal power and maximum remaining energy. Figure 5 shows the flowchart of the proposed method.

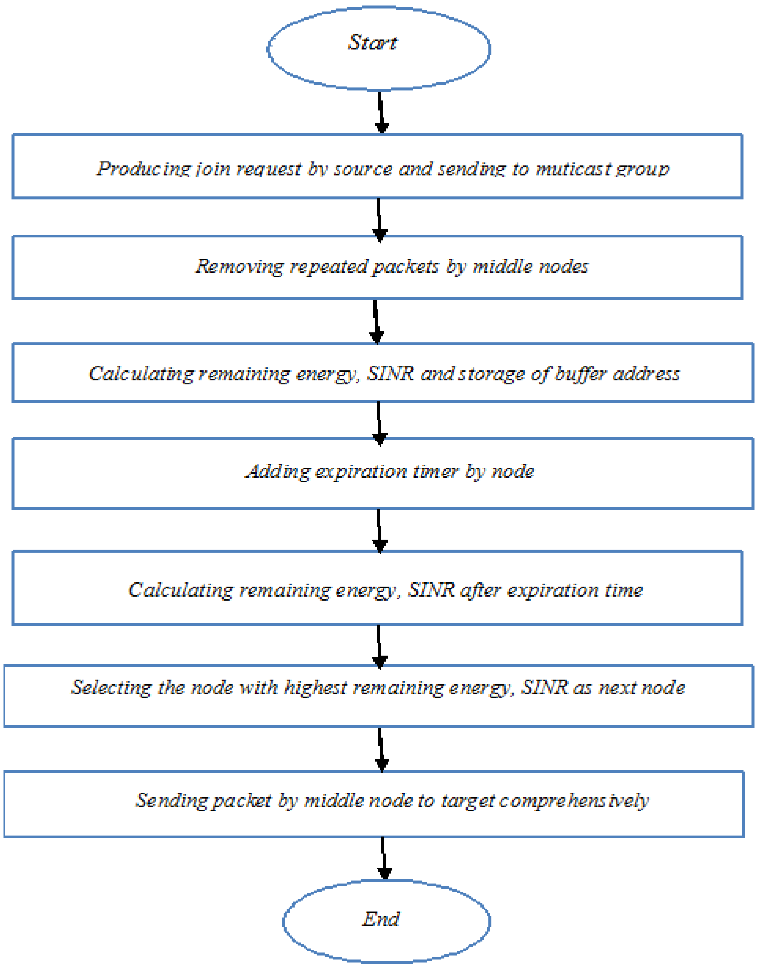


Fig 5. Flochart of the proposed method

A. *Advanced model protocol*

We have used SINR rate as a QoS parameter to improve the power of the network. Our network consists of an undirected graph

G = < ℵ, L>
ℵ = {N1, N2, · · · , Nn}
and a set of nodes L = ℓ1, ℓ2, · · · , ℓm And also a set of links. This network includes one source S and multiple receivers, which are shown by the set D = d1, d2, · · · , dk All receivers are represented by the same multicast address (MCA). It is possible for the receivers to be multiple multicast addresses. There is only one transmitter in a one multicast address. If there is the need to transfer data from S to d | d ∈ D by through middle node of V, we need to estimate the maximum rate of SINR for all input nodes in V at expiration time. The link related to SINR is a "stable link" and is added to the path. In this method, we may have several paths of a node to each other. Each path is a set of stable middle nodes.

VII. EXPERMENTAL DATA AND ANALYSIS

This section includes simulation and evaluation of accurate ODMRP. Compared with IMP-ODMRP with the help of Network Simulator ns-2, we were able to prove, ns is a project simulator, which was introduced in 1989 as a variant of REAL (a network simulator for studying the dynamic behavior of flow and congestion control schemes in packet-switched data networks). We ran two simulations, one for the ODMRP standard, and the other for the proposed IMP-ODMRP method. We repeated the experiments by changing the periods several times up until 200 seconds, and changing the number of nodes from 50 to 100. The simulation parameters are shown in Table 1. The metrics used to evaluate the performance are given below.

Table 1. Simulation Parameters

Simulator	*NS2.34*
Channel type	*Channel/Wireless channel*
Propagation type	*Two ray ground*
Area Simulation	*300m*1500m*
Antenna	*Omni Antenna*
Simulation duration	*200 Sec*
MAC Layer	*802_11*
Traffic Type	*CBR*
Network interface	*Wireless Phy*
Type queue	*Drop Tail*
Number of Node	*50 And 100node*

The following performance metrics have been analyzed: packet delivery ratio, packet loss ratio, average end-to-end delay, packet collision ratio has been regarded as network parameters.

A. *Packet Delivery Ratio (%)*

PDR is the number of packages that are delivered to the destination from the source, divided by the total number of packages in the network. This parameter is also called as success rate of the protocols:

B. *Packet Loss (%)*

Packet drop because of channel congestion, corrupted packets rejected in-transit, faulty networking hardware, faulty

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

network drivers, or normal routing routines (among DSR in mobile ad hoc networks); packet loss can be caused by the black hole attack.

This parameter is calculated using the following formula.

LP = (SP- RP) / SP
LP: Lost Packets
SP: Sent Packet
RP: Received Packet

C. End To End delay

End to end delay sent by node (i) (source node) to packet j, which is temporarily delivered to the destination is as follows:

End_to_End delayij = start_timeij – End_timeij

D. Packet Collision Ratio

As stated in the message, for multicast packets, when recipient nodes attempt to simultaneously acknowledge or repeat their routes, the resulting collision causes packet loss.

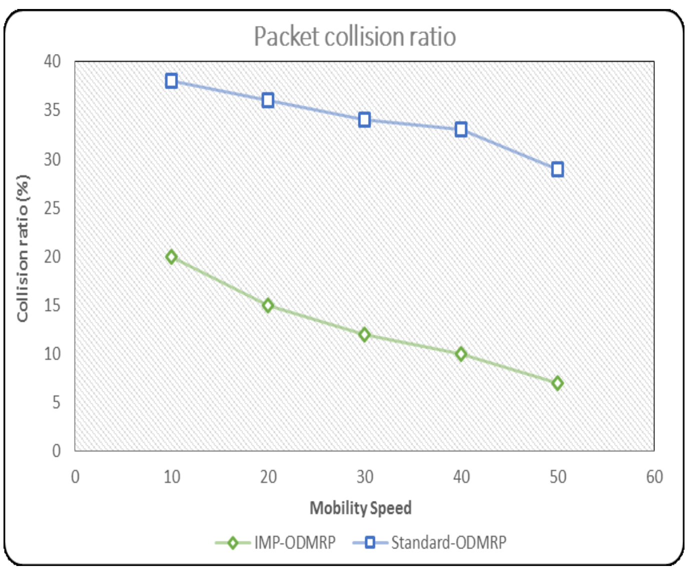


Fig 6. Packet Collision Ratio

Figure 6 shows the proposed method using SINR and the use of sustainable routes to prevent collisions. As expected, the proposed method also exhibited lower collision rates compared to ODMRP.

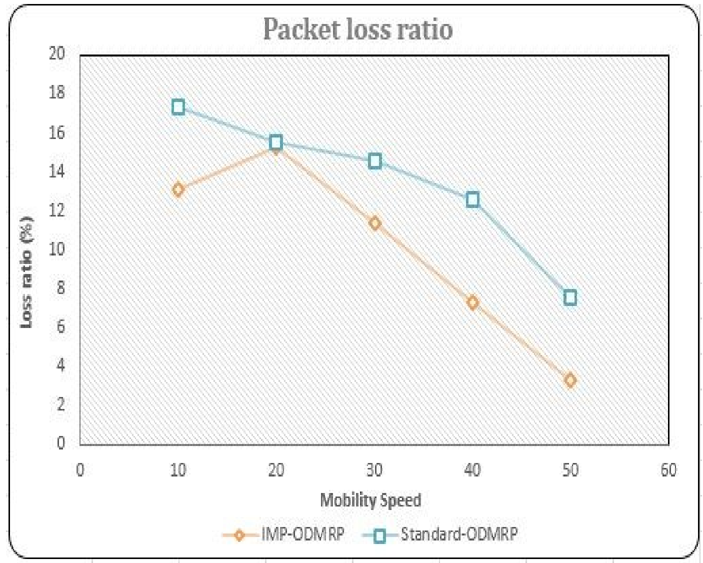


Fig 7. Packet loos Ratio

Figure 7 shows the proposed IMP-ODMRP method compared to ODMRP, at different times reduced the number of missing batches. Due to lower loss ratio depending on the proposed IMP-ODMRP, the route is stable and sends multicast routes using the SINR at each node in conjunction with various transmitters to determine a reliable path.

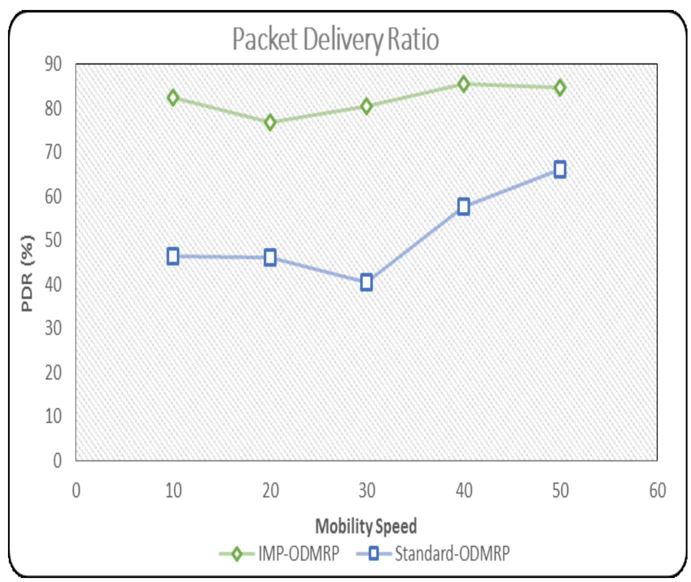


Fig 8. Packet delivery ratio

Figure 8 shows that the (IMP-ODMRP) packet delivery ratio is better than the ODMRP protocol. If the number of broken links is lower, then the number of delivered packets will ultimately increase. The problem is in the method according to the stability of links ODMRP also more than he shows.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

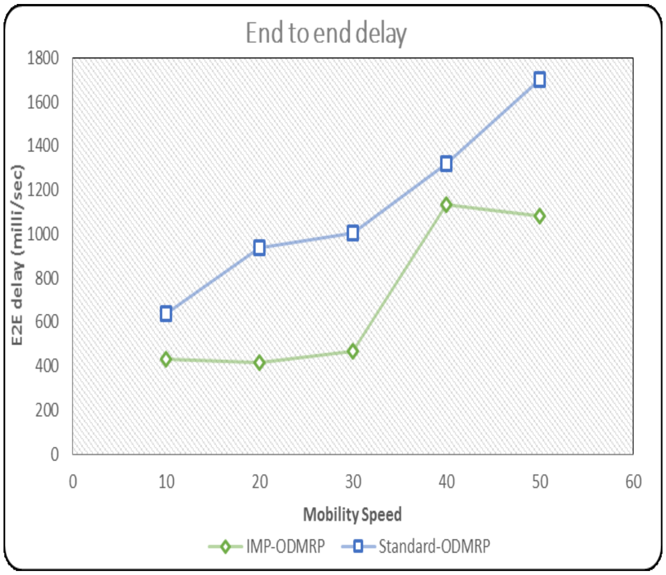


Fig 9. End To End Delay

Figure 9 shows that the (IMP-ODMRP) End-to-end delay is less than the ODMRP protocol. As can be seen, with the increase in node transmission speed, consequently end-to-end delays are reduced, which comparing ODMRP with our proposed method, the added stability of routes in the latter results in even less end-to-end delays.

VIII. CONCLUSION

Today, many researches are conducted regarding multicast routing in mobile ad hoc networks. Multicast routing is used more in this network because of group activities of mobile ad hoc networks. Many protocols have been designed for multicasting in ad hoc networks, which may be tree- based or mesh-based. Qualities like security, functionality, service quality and reliability should be considered in designing multicast protocols. This thesis addressed the routing problem in mobile ad hoc networks and then, multicast protocols were examined in mobile ad hoc networks. We explained new methods of multicast usage and we discussed the multicast protocol ODMRP and its performance. The new proposed method is called IMP-ODMRP. In the proposed method, forward sending of nodes is based on link selection with maximum value of SINR. Link stability provides a more stable path for longer and more persistent network connections. Using the NS-2 simulator, we compared and evaluated our proposed method (IMP-ODMRP) against the ODMRP protocol.

REFRENCE

[1] Al-Hunaity, M., et al. (2007). "A comparative study between various protocols of MANET networks." American Journal of Applied Sciences 4(9): 663-665.
[2] Begdillo, S. J., et al. (2007). "Improving Packet Delivery Ratio in ODMRP with route diversity." IJCSNS 7(12): 146.
[3] Biradar, R. C. and S. S. Manvi (2012). "Review of multicast routing mechanisms in mobile ad hoc networks." Journal of Network and Computer Applications 35(1): 221-239.
[4] Boukerche, A., et al. (2011). "Routing protocols in ad hoc networks: A survey." Computer Networks 55(13): 3032-3080.
[5] Bür, K. and C. Ersoy (2005). "Ad hoc quality of service multicast routing." Computer communications 29(1): 136-148.
[6] Chiang, C. C., et al. (1998). "Forwarding group multicast protocol (FGMP) for mult-ihop, mobile wireless networks." Cluster Computing 1(2): 187-196.
[7] Dhillon, H. and H. Q. Ngo (2005). CQMP: a mesh-based multicast routing protocol with consolidated query packets. Wireless Communications and Networking Conference, 2005 IEEE, IEEE.
[8] Dong, J., et al. (2011). "Secure high-throughput multicast routing in wireless mesh networks." Mobile Computing, IEEE Transactions on 10(5): 653-668.
[9] Dube, R., et al. (1997). "Signal stability-based adaptive routing (SSA) for ad hoc mobile networks." Personal communications, IEEE 4(1): 36-45.
[10] Khani, M., & Ghaffari, A. A New Scheme for Detecting and Preventing Gray Hole Attacks in Mobile Ad Hoc Networks.
[11] Gerharz, M., et al. (2002). Link stability in mobile wireless ad hoc networks. Local Computer Networks, 2002. Proceedings. LCN 2002. 27th Annual IEEE Conference on, IEEE.
[12] Iannone, L. and S. Fdida (2005). Meshdv: A distance vector mobility-tolerant routing protocol for wireless mesh networks. IEEE ICPS Workshop on Multi-hop Ad hoc Networks: from theory to reality (REALMAN).
[13] Khorov, E. and A. Safonov (2010). Multiple Metrics in MANET with End-to-End QoS Support for Unicast and Multicast Traffic. MACOM, Springer.
[14] Lee, S.-J., et al. (2002). "On-demand multicast routing protocol in multi hop wireless mobile networks." Mobile networks and applications 7(6): 441-453.
[15] Moussaoui, A., et al. (2014). "A link-state QoS routing protocol based on link stability for Mobile Ad hoc Networks." Journal of Network and Computer Applications 39: 117-125.
[16] Royer, E. M. and C. E. Perkins (1999). Multicast operation of the ad-hoc on-demand distance vector routing protocol. Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking, ACM.
[17] Santhi, G. and A. Nachiappan (2010). Agent Based Adaptive Multi-constrained Multicast Routing with QoS Guarantees in MANETs. Recent Trends in Networks and Communications, Springer: 200-213.
[18] Ghaffari, A. (2014). An energy efficient routing protocol for wireless sensor networks using A-star algorithm. Journal of applied research and technology, 12(4), 815-822.
[19] Suruliandi, A., and T. Sampradeepraj. "A SURVEY ON MULTICAST ROUTING PROTOCOLS FOR PERFORMANCE EVALUATION IN WIRELESS SENSOR NETWORK." ICTACT Journal on Communication Technology 6.1 (2015).
[20] Toh, C.-K. (1997). "Associativity-based routing for ad hoc mobile networks." Wireless Personal Communications 4(2): 103-139.
[21]Xia, H., et al. (2014). "Applying link stability estimation mechanism to multicast routing in MANETs." Journal of Systems Architecture 60(5): 467-480.
[22] Masoudi, R., & Ghaffari, A. (2016). Software defined networks: A survey. Journal of Network and Computer Applications, 67, 1-25.
[23] Ghaffari, A. (2015). Congestion control mechanisms in wireless sensor networks: A survey. Journal of Network and Computer Applications, 52, 101-115.
[24] Mottaghinia, Z., & Ghaffari, A. (2016). A Unicast Tree-Based Data Gathering Protocol for Delay Tolerant Mobile Sensor Networks. Information Systems & Telecommunication, 59.
[25] Ghaffari, A. Real-time routing algorithm for mobile ad hoc networks using reinforcement learning and heuristic algorithms. Wireless Networks, 1-12.
[26] Alizadeh, S., & Ghaffari, A. (2010, July). An Energy-efficient hirerchical Clustering protocole in wireless sensor networks. In Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on (Vol. 9, pp. 413-418). IEEE.
[27] KeyKhosravi, D., Ghaffari, A., Hosseinalipour, A., & Khasragi, B. A. (2010). New Clustering Protocol to Decrease Probability Failure Nodes and Increasing the Lifetime in WSNs. Int. J. Adv. Comp. Techn., 2(2), 117-121.

AUTHORS PROFILE

Ebrahim Asadi received his B.Sc. in computer engineering from AHRAR University, Rasht, Iran, in 2011 and his M.Sc. in computer engineering from Azad university of Shabestar branch,Tabriz, Iran, in 2015. His research interests Mobile Ad-Hoc Networks.

Ali Ghaffari received his B.Sc, M.Sc and Ph.D degrees in computer engineering from the University of Tehran and IAUT(Islamic Azad University), TEHRAN, IRAN in 1994, 2002 and 2011 respectively. As an assistant professor of computer engineering at Islamic Azad University, Tabriz branch, IRAN, his research interests are mainly in the field of wired and wireless networks, Wireless Sensor Networks (WSNs) ,Mobile Ad Hoc Networks (MANETs),Vehicular Ad Hoc Networks (VANETs),networks security and Quality of Service (QoS). He has published more than 50 international conference and reviewed journal papers.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 6, June 2016

A Survey on Human Social Phenomena inspired Algorithms

Thanh Tung Khuat, My Hanh Le

***Abstract*—The problem of seeking the optimal solution in the field of science and engineering has been becoming complex and challenging due to the explosion of dimensions and the interdependence of variables. Over the past few decades, a variety of new concepts, techniques and computational applications inspired from nature have been proposed and used to deal with a wide range of optimization problems in diverse fields. Many of nature-inspired algorithms generate high-quality solutions for real-world optimization tasks. Nevertheless, the majority of these methods are inspired by either biological phenomena or social behaviors of mainly animals and insects. There are few works relied on social phenomena of human being used to form optimization algorithms. This paper aims at presenting an adequate review of most predominant and successful groups of optimization approaches based on human social phenomena.**

***Index Terms*—Human Social Phenomena, Society Civilization Algorithm, Cultural Algorithms. Teaching-learning-based Optimization, Social Learning Algorithm, Alliance Formation based Algorithms, Social Emotional Optimization Algorithm, Social Labeling.**

I. Introduction

THE field of nature-inspired computing has been becoming prevalent in recent years. Nature-inspired algorithms have achieved the enormous success when applied to real-world optimization problems. The majority of these algorithms inspire by the evolutionary principles of Darwin [1], the social and cognitive behavior of animals [2] and physical phenomena [3]. These sources of inspiration are able to tackle the optimization problems of almost all areas including wireless sensor networks, computer networks, control systems, image processing, data mining, parallel processing, robotics, and biomedical engineering, *etc*.

It can be seen that most of the natural algorithms in the optimization domain are inspired by either biological phenomena or social behaviors of mainly animals and insects. There are only a few studies based on social phenomena in human societies to form optimization algorithms. This paper produces an overview of most common and successful groups of the state-of-the-art of human social phenomena inspired algorithms.

The computational power of the algorithms inspired by social phenomena is the combination of the richness and complexity of the social behaviors with interaction among individuals in the population. Human interactions cannot be explained only by means of biological information and they are more likely tend to demonstrate a high level of diversity. Even if there is not an optimizing principle behind social interactions, they also result in robust social structures that may even show a high level of stability. Human society is a complex group which is more effective than other animal categories. Hence, if one algorithm simulates the human society, the efficiency and effectiveness may be more robust than other swarm intelligent mechanisms which are inspired by other animal groups. These characteristics enable social phenomena to become a valuable source of inspiration for algorithms [4]. There are some human social phenomena forming the idea for the optimization algorithms such as leadership and influence from prominent counterparts, teaching and learning, alliance formation, social labeling of individuals, and social emotion. These sources of inspiration have been become the main idea for the appearance of optimization algorithms in the human society-inspired computing fields.

The rest of this paper is organized as follows. Section II represents leadership based algorithms including society civilization and cultural algorithms. Section III shows algorithms based on teaching and learning behaviors in the human society. Alliance Formation is also a source of inspiration for optimization algorithms as shown in Section IV. Section V describes a social emotional optimization algorithm while Section VI presents about optimization algorithm through Social Labeling. Finally, section VII shows the conclusion of this survey.

II. Leadership Based Algorithms

Human beings are social animals and living together in large groups naturally meant that people needed to use various roles and accomplish different groups. In order to give structure to society and help society grow and develop, people were divided into leaders and followers. Leaders play a vital role and affect the rest of the society, and the society with the absence of leadership will result in chaos. Therefore, leadership has become the source of inspiration for optimization algorithms.

Thanh Tung Khuat is now with the Department of Information Technology, The University of Danang - University of Science and Technolgoy, Danang, Vietnam (e-mail: thanhtung09t2@gmail.com).

My Hanh Le is with the the Department of Information Technology, The University of Danang - University of Science and Technolgoy, Danang, Vietnam (e-mail: ltmhanh@dut.udn.vn).

A. Society Civilization Algorithm

Society civilization algorithm (SCA) [5] is based on the social phenomena including migration, leadership and cooperation by leaders sharing their knowledge, which is their position in search space, to the remainder of individuals. In this algorithm, candidate solutions which are distributed in clusters resembling societies explore search space at the beginning stage. The best solutions in each society bias the search toward them, so the leaders influence their counterparts to follow them. Some leaders may migrate to other regions and therefore their counterparts in former societies travel with them. The set of all societies is considered as the civilization, in which all individuals might interact by means of their leaders. The main steps of the algorithm are presented as in Fig. 1.

Unlike evolutionary algorithms, where only the better performing individuals are enabled mate to produce children, the SCA refines the efficiency of all individuals in every society, either through an intra or intersociety information exchange [5]. The algorithm also holds on parametrically unique solutions across the civilization, which at the end corresponds to a set of elite near optimal individuals. The maintaining process of unique individuals requests additional computations but it supplies a room for diverse solutions to exist and evolve. In this algorithm, authors adopted a leader centric operator that results in search around the leader to move an individual toward the location of leaders.

1. t ← 0
2. Generate individuals representing a civilization of *N* individual: $C(t) = \{I_1,\ldots,I_N\}$ uniformly in the parametric space.
3. Assess individuals by calculating the objective function as well as constraints
4. Build societies *S*(t) by creating *K*(t) clusters from *C*(t). Clusters are able to be constructed by any cluster analysis technique.
5. Identify leaders in each society
6. Move individuals in each society toward the location of its nearest leaders
7. Determine leaders in the civilization *C*(t) from the leaders of the societies *S*(t)
8. Move all society leaders toward the location of civilization leaders
9. t ← t + 1
10. If the stop condition is not satisfied then go to step 3

Fig. 1. Society Civilization Algorithm

Through the experimental results, authors pointed out that SCA exhibited a very fast convergence in all the examples. This feature is particularly attractive because it would mean attaining comparable solutions with fewer function evaluations. SCA is applied to optimize the economic dispatch with multiple minima, a well-known problem in electric power systems operation, and results are promising as effectiveness is comparable to those from mathematical programming, with less computational effort.

B. Cultural Algorithms

A unique aspect characterizing all human societies is the concept of culture which is based on learning from experienced individuals and guidance from the best individuals to the rest of ones. Based on how cultures gather information to solve the problems, Reynolds [6] proposed an approach called cultural algorithms (CA). Such algorithms are a vehicle for modeling social evolution and learning. The key idea of CA is to divide the process of learning and information retrieving into three phases. First, a coarse-grained phase is set up with the expectation to grasp a general idea of the problem so as to specify regions to explore. Then, there is a fine grained phase and finally a phase comes into action when the search process is stagnated. This is an abstraction of how cultures learn to deal with their problems by methods of a space belief.

The model of Cultural Algorithms is an expression of the THINK model of Renfrew [7] including a belief space containing individual and group mappa and a trait-based population space. Mappa are viewed as being subsets of the belief space. In the model of CA, each individual is shown in terms of a set of traits or behaviors and a mappa of its experience. Traits can be changed and exchanged between individuals meanwhile individual mappa are able to be merged and modified to form "group mappa". At any given time step in the model, there are a set of individuals in the population space and the performance of each individual is evaluated as well as each individual will create a generalized map of its experience during that time period. A mappa of individual then is merged with currently existing group mappa in the belief space if the conditions for merging operators are met. When mappa are merged, the performances of individuals associated with them are combined. If the combined performance of a mappa is less than the acceptable level then that mappa is discarded. The current state of the belief space can be used to change the performance of individuals in the population. The population is then used to produce a new population via the selection of parent individuals for the next generation. These parents are utilized to evolve a new population by applying different modification operators. The main steps of the CA are presented in Fig. 2.

1. Initialize population with *k* individuals
2. Evaluate individuals in the population
3. Initialize the belief space
4. **repeat**
5. Apply mutation to generate *t* offspring
6. Evaluate each offspring
7. Compute the relative performance of each individual by the method of random mutations
8. Select *q* individuals with the largest number of victories to form the new generation
9. Add the non-dominated individuals to an external memory
10. Modify the belief space with individuals in external memory
11. **until** the stop condition is satisfied

Fig. 2. Cultural Algorithms

III. Teaching And Learning Based Algorithms

Teaching and learning are two in the most popular activities of human society and also are a source of inspiration for algorithms.

A. Teaching-Learning based Optimization Algorithm

Rao *et al.* [8] introduced teaching-learning-based optimization (TLBO) method based on the philosophy of the teaching-learning process. For the TLBO, the population is considered as a group of learners or a class of learners. The

search process contains two phases: Teacher phase and Learner phase.

1) Teacher phase

In the teacher phase, learners acquire knowledge from a teacher. In the entire population, the best solution is considered as the teacher ($\vec{X}_{teacher}$). The teacher tries to improve the results of other individuals ($\vec{X}_i$) by increasing the average result of the classroom ($\vec{X}_{mean}$) towards his/her level [8]. The solution is updated according to the difference between the existing and the new mean given by Eq. (1).

$$\vec{X}_{new}=\vec{X}_i+r_i*(\vec{X}_{teacher}-T_f*\vec{X}_{mean}) \qquad (1)$$

where T_f is a teaching factor that decides the value of mean to be changed, and r_i is a random number in the range of [0, 1]. The value of T_f can be either 1 or 2, which is again a heuristic step. Moreover, $\vec{X}_{new}$ and $\vec{X}_i$ are the new and existing solutions of the i^{th} learner, respectively.

2) Learner phase

In the learner phase, learners try to increase their knowledge by interacting with others. A learner interacts randomly with other learners with the help of group discussions, presentations, formal communications, *etc*. [8]. A learner learns something new if another learner has more knowledge than him or her. The modification of the learner is represented as Eqs. (2)-(3).

$$\vec{X}_{new} = \vec{X}_i + r_i * (\vec{X}_j - \vec{X}_k) \quad \text{if } f(\vec{X}_j) < f(\vec{X}_k) \qquad (2)$$

$$\vec{X}_{new} = \vec{X}_i + r_i * (\vec{X}_k - \vec{X}_j) \quad \text{if } f(\vec{X}_k) < f(\vec{X}_j) \qquad (3)$$

where, $\vec{X}_k$ and $\vec{X}_j$ ($j \neq k$) are two students chosen randomly in the population, and f is the fitness function. If the new solution $\vec{X}_{new}$ is better, it is accepted in the population. The algorithm will continue until the termination condition is met. The main steps of the TLBO are presented in Fig. 3.

```
Initialize the population of learners including N individuals
Evaluate the quality of each learners in the class
Choose the best learner as X_teacher and compute the mean X_mean of all learners
while (stopping condition is not met)
   for all learners
          Update all learners following the Eq. (1)
   end for
   Evaluate the new learners
   Accept the new solution if it is better than the old one
   for all learners
          Randomly select two other learners being different from it
          Update the learner according to the Eqs. (2)-(3)
   end for
   Accept the new solution if it is better than the old one
   Update the teacher and the mean
end
```

Fig. 3. Teaching-Learning based Optimization Algorithm

The performance of TLBO is tested by experimenting with different benchmark problems with various characteristics. The results indicated the better performance of TLBO over other nature-inspired optimization methods for the constrained benchmark functions and mechanical design problems taken into account. In addition, TLBO shows a better performance with less computational effort for large scale problems, *i.e.* problems of a high dimensionality.

B. Social Learning Algorithm

Social learning theory presents how people learn in a social context in which the adjustment of behaviors can occur either from direct experience or by observing other people. The observational learning begins with an attention process. People remember the details of their exemplary behavior with a retention process and practice to reproduce the behavior with a reproduction process. Reinforcement plays a crucial role to distinguish learning from simply imitating the others. Based on Bandura's Social Learning Theory [9] which pointed out the social learning behavior of humans being a high level of intelligence in nature, Jiao *et al.* [10] proposed a novel evolutionary computation algorithm called Social Learning Algorithm (SLA) which emulates the social intelligence of humans in computers as shown in Fig. 4. After the initialization, SLA carries out an iteration process in which the members conduct 'attention', 'reproduction', 'reinforcement', and 'motivation' operators repeatedly. The four operators are like to those in the process of observational learning in social learning theory.

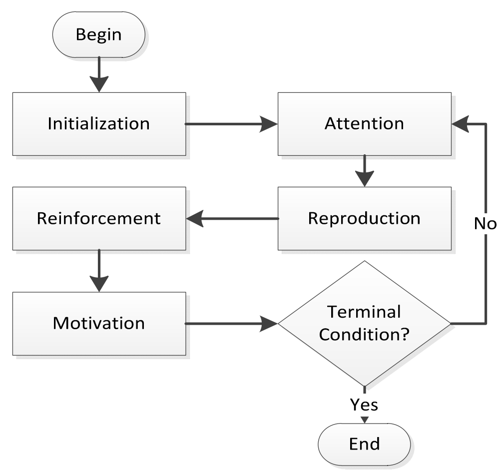


Fig. 4. Flowchart of the Social Learning Algorithm

Like the other population-based optimization approaches, SLA maintains a population of individuals, specifically, a social group of people. Each member i in the group is assigned with a behavioral pattern vector $x_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,D}]$, where D is the dimensionality of the problem.

Attention operator identifies whose and which characteristics capture attention in the social group following the scores. The entire society is divided equally into two segments: upper society ***US*** and lower society ***LS***. For each dimension $d \in \{1, 2, \ldots, D\}$ of the society, students' t-test is conducted to compare the values of ***US*** and ***LS*** on this dimension and the calculated statistical t-value is recorded as $t(d)$. Let $AT = |t(r)|$ with r is a randomly selected dimension index. Next, mark each dimension of the society with an attention symbol $\Gamma(d)$ as Eq. (4).

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$\Gamma(d)=\begin{cases}\gg & if\ t(d)\geq AT\\ \approx & if\ -AT<t(d)<AT\\ \ll & if\ t(d)\leq -AT\end{cases} \quad (4)$$

The reproduction step is performed after attention to construct new behavior vectors for all members by imitation. Each member i in the entire society generates a new behavior vector $x'_i = [x'_{i,1}, x'_{i,2},\ldots, x'_{i,D}]$. For each dimension d, if $\Gamma(d) = \gg$ or $\ll$ then $x_{i,d} = x_{ri,d}$ where ri is a random index of members in ***US***. For dimensions are marked with '≈', x'_i maintains the values of x_i or explores the entire society according to the Eq. (5).

$$x'_{i,d}=\begin{cases}x_{e,d} & if\ r_1<PI\\ rand(l_d,u_d) & otherwise\ if\ r_2<PR\\ x_{i,d} & otherwise\end{cases} \quad (5)$$

where e is the index of a member randomly selected from the entire society; l_d and u_d are the lower and upper bounds of the variable on dimension d; PI is the probability of imitation; PR stands for the probability of randomization; r_1 and r_2 are random numbers in the range of [0, 1].

The further reinforcement enhances the learnt behavior with positive reward or negative punishment. In SLA, positive reinforcement and negative punishment are conducted on dimensions marked with '<<' and '>>' as Eq. (6).

$$x'_{i,d}=\begin{cases}x'_{i,d}+\Delta_{i,d} & if\ \Gamma(d)=\gg\\ x'_{i,d}-\Delta_{i,d} & if\ \Gamma(d)=\ll\end{cases} \quad (6)$$

where $\Delta_{i,d} = rand(0,1) * |x_{ri,d} - x_{i,d}|$

The motivation process activates the new behavior vectors with incentives. In SLA, after reproduction and reinforcement, the learnt behavior vector x'_i of each member is assessed. Only when x'_i achieves a higher score than the current behavior x_i, the member updates his behavior vector.

IV. Alliance Formation Based Algorithms

Another social phenomenon which is the source of inspiration for optimization algorithms is Alliance Formation. An alliance is relatively stable group of agents being perhaps people, political parties, software programs, robots or firms that work toward a common purpose. Several alliance formation algorithms have been introduced in the context of multi-agent systems [11], [12], [13]. The key idea is to maximize the sum of the payoffs to all the alliances by determining the optimal combination of alliances and the division of agents into these alliances. Each agent has a set of tasks that may or may not be similar to the tasks assigned to other agents. A basic idea of the algorithms based on Alliance Formation is shown in Fig. 5.

Alliance formation algorithms may be grouped into two classes. The first class includes dynamic programming algorithms that assure to seek the optimal alliance. The second class is heuristic-based algorithms that do not guarantee finding the optimal, but reach solutions very fast. In the coalition formation in human societies inspired algorithms, each agent i is given a variable quantifying its strength of character determining its predisposition to construct complete new alliances, C_i. A second variable G_i defines its attraction to achieve a profit within that alliance. Eventually, a third variable R_i is defined that states its reluctance to abandon its current alliance to join another one. Agents with a high C_i will send invitations to form new coalitions more often than those with lower predisposition. When an agent i that is part of a coalition A_{cur} is invited to join another alliance A_n, its decision is based not only on the benefit x_i^n it will receive but also on the individual parameters that define its personality. The gains from both the new and actual coalition, $S(A_n)$ and $S(A_{cur})$ are determined as Eqs. (7)-(8).

$$S(A_{\mathrm{cur}})=G_i*\frac{x_i^{cur}}{x_i^{cur}+x_i^n}+R_i*(1-\mathrm{sb}_{cur}) \quad (7)$$

$$S(A_n)=G_i*\frac{x_i^{n}}{x_i^{cur}+x_i^n}+R_i*(1-\mathrm{sb}_{n}) \quad (8)$$

where sb_{cur} is a parameter summarizing the stability of the coalition that consists of agent i, and sb_n is the stability of the new coalition. Therefore, if $S(A_{cur}) \leq S(A_n)$, the agent decides to abandon alliance A_{cur} and to join the alliance A_n.

1. Create a list of initial alliance structures with q agents $S(q)$
2. **while** S(q) is not empty
3. Contact agents $A_i \in S(q)$
4. Assess the benefit of joining A_i in an alliance based on preferences and constraints.
5. Extract q from $S(q)$ and share subtasks
6. If contacted by agent A_k, substract q as well as the common tasks
7. **end**

Fig. 5. Alliance Formation based Algorithms

V. Social Emotional Optimization Algorithm

In human society, all people do their work hardly to increase their society status. In order to achieve this objective, people will try their bests to find the path so that higher rewards can be obtained from society. This idea is inspired to create a social emotional optimization algorithm (SEOA) [14]. In SEOA, each individual represents one person; while all points in the problem space erects the status society. In this virtual world, all individuals aim to find the higher social status. Hence, they will communicate through cooperation and competition to raise personal status, and the one with highest score will win and output as the final solution. In SEOA, all individuals' decisions are affected by one constant emotion selection strategy. Nevertheless, this strategy is able to supply a wrong search selection because of some randomness omitted.

In the first step, all individuals' emotion indexes are set to 1, and all individuals' emotion indexes are the largest value, thus, they think their behavior in this iteration is right, and select the next behavior as Eq. (9).

$$\overrightarrow{x_j}(1)=\overrightarrow{x_j}(0)\oplus \mathrm{M}_1 \quad (9)$$

where $\overrightarrow{x_j}(1)$ shows the degree of the j^{th} individual in the initialization period and its fitness value is denoted as the social status. Symbol $\oplus$ means the operation and we usually take care of addition operation +. Because the belief index of j is 1, the next behavior movement phase M_1 is specified by Eq. (10).

$$M_1 = -k_1 * rand_1 * \sum_{i=1}^{L} (\overrightarrow{x_i}(0) - \overrightarrow{x_j}(0)) \tag{10}$$

where k_1 is a parameter used to control the emotion changing size, $rand_1$ is a random number with uniform distribution in the range of [0, 1]. Total L individuals are chosen whose status values are the worst to provide a reminder for individual j to avoid the wrong behaviors.

In the t^{th} generation, if individual j does not attain one better society status value than all previous values, the emotional index of individual j is decreased as Eq. (11).

$$EM_j(t+1) = EM_j(t) - \Delta \tag{11}$$

where Δ is a predefined value and is usually taken from experimental tests. If individual j is rewarded a new status value which is the best one among all iterations, then:

$$EM_j(t+1) = 1.0 \tag{12}$$

If $EM_j(t+1) < 0$ then $EM_j(t+1) = 0$.

To simulate the behavior of human, a set of three kinds of manners are defined $\{M_2, M_3, M_4\}$. Because the emotion impacts on the behavior, the next behavior will be changed according to the following three rules as Eqs. (13)-(15).

If $EM_j(t+1) < T_1$ then:

$$\overrightarrow{x_j}(t+1) = \overrightarrow{x_j}(t) + M_2 \tag{13}$$

If $T_1 \le EM_j(t+1) < T_2$ then:

$$\overrightarrow{x_j}(t+1) = \overrightarrow{x_j}(t) + M_3 \tag{14}$$

Otherwise,

$$\overrightarrow{x_j}(t+1) = \overrightarrow{x_j}(t) + M_4 \tag{15}$$

Parameters T_1 and T_2 are two thresholds aiming to curb the different behavior manner. Manners are updated as Eq. (16).

$$M_2 = k_2 * rand_2 * (\overrightarrow{\mathrm{St}}_{best}(t) - \vec{x}_j(t)) \tag{16}$$

where $\overrightarrow{\mathrm{St}}_{best}(t)$ shows the best society status degree obtained from all people previously and

$$\overrightarrow{\mathrm{St}}_{best}(t) = \arg\min_{s}\{\mathrm{f}(\overrightarrow{x_s}(\mathrm{h}) \mid 1 \le \mathrm{h} \le \mathrm{t})\} \tag{17}$$

For M_3, we have

$$M_3 = k_3 * rand_3 * (\overrightarrow{x_{j\,best}}(t) - \vec{x}_j(t)) + M_2 + M_1 \tag{18}$$

where $\overrightarrow{x_{j\,best}}(t)$ denotes the best status value obtained by individual j previously, and is defined by:

$$\overrightarrow{x_{j\,best}}(t) = \arg\min\{\mathrm{f}(\overrightarrow{x_j}(\mathrm{h}) \mid 1 \le \mathrm{h} \le \mathrm{t})\} \tag{19}$$

For M_4, we obtain:

$$M_4 = k_3 * rand_3 * (\overrightarrow{x_{j\,best}}(t) - \vec{x}_j(t)) + M_1 \tag{20}$$

The details of social emotion optimization are presented as in Fig. 6.

1. Initialize all individuals with randomly initial position of individuals
2. Calculate the fitness value of each individual according to the objective function
3. For individual j, determining the value $\overrightarrow{x_{j\,best}}(0)$
4. For all population, determining the value $\overrightarrow{\mathrm{St}}_{best}(0)$
5. Determine the emotional index according to Eqs. (13)-(15) in which three emotion cases are determined for each individual
6. Identify the decision with Eqs. (16)-(20)
7. Making mutation operation
8. If stopping criteria is met, output the best solution; otherwise, goto step 3

Fig. 6. Social Emotional Optimization Algorithm

Strategy mentioned above is able to supply a wrong search selection because of some randomness omitted. To enhance the performance, Cui *et al*. [15] proposed three different random emotional selection strategies. Simulation results indicated that SEOA with Gauss distribution is more effective than the standard version of algorithm.

VI. Optimization Through Social Labeling

In a society, categorization of individuals is a popular phenomenon. The label or tag assigned to an individual might be the result of ignorance, prejudices, or true facts and it affects the way other individuals interact with him/her by inducing a positive or negative feeling. Several algorithms have been inspired by such social tags.

In [16], Hales and Edmonds proposed a method in order to improve cooperation in Peer to Peer (P2P) networks based on the evaluation among peers resulting in each individual being assigned a tag. This algorithm helps to achieve cooperation in P2P networks effectively. The details of this algorithm are shown in Fig. 7.

```
1. while (the number of generations is not met)
2.       for each agent i in the population
3.             Select a game partner agent j with a similar tag
4.             Peers i and j interact through their strategies and get payoff
5.       end for
6.       Reproduce agents proportionally to their payoff
7.       Mutate tags and strategies of each reproduced agent
8. end while
```

Fig. 7. Optimization Algorithm using Social Labeling for P2P Networks

In this algorithm, each peer is allocated a strategy that states its behavior for interacting with other agents. Two agents are involved in a situation where they have the option to cooperate (C) with each other or to defect (D). Each agent achieves a payoff as a function of its action and the action of the other peers as follows:

$$S < P < R < T$$

$$T + S < 2R$$

where T is the payoff an agent gets if it defects and the other

cooperates, R is the payoff when both agents cooperate, P is the payoff when both of them defect, and S is the payoff an agent getting if it cooperates and the other defects.

Optimization is obtained by selecting a more convenient counterpart (step 3 in Fig. 7) to have a significant level of cooperation [16]. When tags are eliminated, the achieved network is highly inefficient, as the peers do not obtain the desired resource due to the lack of cooperative agents.

Another algorithm relied on social tags was presented in [17]. A mining newsgroup algorithm leads to the taxonomy of people in two opposite camps over a discussion issue. The algorithm is based on the assumption that people respond more frequently to messages that consist of ideas they do not agree with. Through their responses, people get tagged and this tag is considered to define the topology of a bipartite network in which each vertex shows a participant and edges present responses between participants. This algorithm finds a partition of the vertices into two sets: F and A, one presenting participants in favor and the other showing users against the discussion issue.

VII. CONCLUSION

This paper presented the summary of algorithms inspired by human social phenomena. Many human activities may become the source of inspiration for algorithms such as leadership, alliance formation, society status, teaching, learning and categorization of individuals in the society. Human society-inspired algorithms can apply for solving many optimization problems in the real-world involving computer networks, wireless sensor networks, robotics, control systems, parallel processing, biomedical engineering, and image processing. Nonetheless, this work is one of the very few survey conducted on social phenomena based optimization methods.

Most of the algorithms relied on social phenomena give the outcomes equivalent to ones inspired by social behaviors of animals and insects. The results shown in the literatures are encouraging, but many ideas and metaphors from human social phenomena are still waiting for further studies.

REFERENCES

[1] C. Darwin, *The origin of species*, 6th ed.: Cambridge University Press, 2009.

[2] J. Kennedy, R. C. Eberhart, and Y. Shi, *Swarm Intelligence*. San Francisco: Morgan Kaufmann, 2001.

[3] A. Kaveh, *Advances in Metaheuristic Algorithms for Optimal Design of Structures*: Springer, 2014.

[4] R. Chiong, *Nature-Inspired Algorithms for Optimisation* vol. Vol. 193: Springer, 2008.

[5] T. Ray and M. Liew, "Society and civilization: An optimization algorithm based on the simulation of social behavior," *IEEE Transactions on Evolutionary Computation,* vol. 7, pp. 386–396, 2004.

[6] R. Reynolds, "An introduction to cultural algorithms," in *Proceedings of the 3rd Annual Conference on Evolutionary Programming*, 1994, pp. 131–139.

[7] A. C. Renfrew, "Dynamic modeling in Archaeology: What, When and Where?," in *Dynamical Modeling and the Study of Change in Archaeology*, S. E. V. d. Leeuw, Editor, Endinburgh, Scotland: Endinburgh University Press, 1994.

[8] R. V. Rao, V. J. Savsani, and D. P. Vakharia, "Teaching–Learning-based optimization: A novel method for constrained mechanical design optimization problems," *Computer-Aided Design,* vol. 43, pp. 303-315, 2011.

[9] A. Bandura, *Social Learning Theory*. UK: Oxford, 1977.

[10] G. Y. Jiao, Z. Jun, and L. Yun, "From the social learning theory to a social learning algorithm for global optimization," in *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2014, pp. 222-227.

[11] O. Shehory and S. Kraus, "Task allocation via coalition formation among autonomous agents," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1995, pp. 655–661.

[12] K. Kijima, "Agent-based Simulation of Alliance Formation and its Stability Analysis: Application to Aviation Industry," in *PACIS 2003 Proceedings*, 2003, p. 81.

[13] V. Mashkov, "Restricted alliance and coalitions formation," in *Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, 2004, pp. 329-332.

[14] Z. H. Cui, Z. Z. Shi, and J. C. Zeng, "Using social emotional optimization algorithm to direct orbits of chaotic systems," in *Proceedings of 2010 International Conference on Computational Aspects of Social Networks (CASoN2010)*, 2010, pp. 389-395.

[15] Z. Cui, Y. Xu, and J. Zeng, "Social Emotional Optimization Algorithm with Random Emotional Selection Strategy," in *Theory and New Applications of Swarm Intelligence*, ed: InTech, 2012, pp. 33-50.

[16] D. Hales and B. Edmonds, "Applying a socially inspired technique (tags) to improve cooperation in P2P networks," *IEEE Transactions on Systems, Man and Cybernetics* vol. 35, pp. 385-395, 2005.

[17] R. Agrawal, S. Rajagopalan, R. Srikant, and Y. Xu, "Mining newsgroups using networks arising from social behavior," in *Proceedings of the 12th International Conference on WWW*, 2003, pp. 529–535.

Thanh Tung Khuat completed the B.S degree in Software Engineering from University of Science and Technology, Danang, Vietnam, in 2014. Currently, he is participating in the research team at DATIC Laboratory, University of Science and Technology, Danang. His research interests include software engineering, software testing, evolutionary computation, Intelligent Optimization Techniques and Applications in software engineering.

My Hanh Le is currently a lecturer of the Information Technology Faculty, University of Science and Technology, Danang, Vietnam. She gained M.Sc. degree in 2004 and Ph.D. degree in Computer Science from The University of Danang in 2016. Her research interests are about software testing, nature-inspired algorithms and more generally application of heuristic techniques to problems in software engineering.

Mammogram Classification Using Selected GLCM Features and Random Forest Classifier

Vibhav Prakash Singh*, Ayush Srivastava, Devang Kulshreshtha, Arpit Chaudhary, Rajeev Srivastava
Department of Computer Science & Engineering, Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh-221005, India

*Corresponding Author

***Abstract*- Early diagnosis of breast cancer can improve the survival rate by detecting the cancer at initial stage. Mammogram is a low dose X-ray image of the breast region, used to diagnose the breast cancer at early stage. In this paper, an efficient computer added diagnosis (CAD) system is proposed, automatically detects the normal and abnormal images of mammogram. The proposed pre-processing steps include, cropping of mammograms (for avoiding the pectoral muscle, unwanted tags) and suppression of Gaussian noise. Further, gray level co-occurrence matrix (GLCM) based statistical texture feature from different distances of neighboring and angles are extracted. Furthermore, most relevant features are also examined using AdaBoost feature selection method. Finally, normal and abnormal mammograms are classified using Random forest (RF) classifier. Experiments on benchmark mammography image analysis society (MIAS) database confirm the effectiveness of this work**.

Keywords-CAD; Mammography; GLCM features; Feature selection; Random forest classifier.

I. INTRODUCTION

Breast cancer is a type of cancer which develops from a mutated gene and forms in tissues of the breast, usually the ducts and globules. It occurs in both men and female but male breast cancer is rare to find. It is the second most common cancer in women after skin cancer [9]. The risk of developing a breast cancer in women is very high after 60 years. Generally, breast cancer is diagnosed in the age between 45 to 64 years. Only 14 % of cases are detected under the age of 45 years, and 37 % cases are diagnosed above the age of 65 years [18]. Breast cancer is the most common and leading cause of death in women, and detection at early stages means to find breast cancer in women who do not show any symptoms. The most important screening test for breast cancer detection is the mammography. Mammography is a medical imaging process, uses low-dose X-ray system to see inside the breasts [19]. The images generated through the process of mammography is known as mammograms. In medical imaging areas, developing computer-aided diagnosis (CAD) and detection is a very growing research area from the last two decades [1, 16]. CAD schemes have been developed for variety of medical images just like lung computerized tomography (CT) images, mammography, chronic obstructive pulmonary disease, MRI, pulmonary embolism, CT colonography, and other pathology images. Among these different types of CAD schemes, CAD of mammograms is the most mature one [17]. It reduces the mortality among women due by identifying the tumor in initial stage using machine learning approaches and content based image retrieval (CBIR) [14]. There is increasing interest in the use of CBIR techniques to diagnose the stage of breast cancer by identifying similar past cases. CBIR is the process of retrieving desired images from a large collection based on visual similarity of current query image. Mammogram image classification categories the images into different classes, can be treated as a pre-processing step and applied before computing the similarity measure to speed up the searching and retrieval performances of a CBIR system[12]. Mammogram image classifications can also help for the early diagnosis of breast cancer by detecting the cancer in mammograms. The success of accurately classifying mammograms depends on what features that are extracted from mammograms and fed into the learning model. Hence, we need to transform processed images into features that are better represent the task of classification. As we know that mammogram visual appearance are much closed to texture based appearance [3, 6, 7]. So in this study, we have used grey level co-occurrence matrix (GLCM) based texture features [4], capable for effective classification for normal and abnormal mammogram. Further, feature selection is a very relevant step in creating an accurate predictive classifier. It can be used to identify and remove irrelevant and redundant features from data that do not contribute in discriminating among normal and abnormal mammograms or may in fact decrease the accuracy of the model. Less attributes are also desirable because they reduce the speed and complexity of classifier training and execution. Moreover, the normal and abnormal images contain some similar characteristics. Hence, it is practical to eliminate similar features between normal and abnormal mammograms.

In this work, we classify or detect the cancerous and non-cancerous images using random forest classifier and GLCM based informative texture features, where informative features are selected using AdaBoost feature selection method. In the next section, we demonstrate the proposed framework, start with pre-processing, followed by feature extraction, feature selection and image classification. Results analyses are presented in Section III, and finally Section IV presents the conclusion of the work.

II. METHODS AND MODELS

In this section, we explore the proposed work, based on selective GLCM feature sets. For better classification, random forest classifier is introduced, which are good enough to reduce the effect of over-fitting and classify images effectively. This work presented in this paper is divided into 4 sections (A, B, C and D), where section A gives the detail about region of interest cropping and filtering, section B demonstrate details and extraction procedure of used features. Section C gives the complete algorithm of AdaBoost feature selection method, finally section D sheds some light on mammogram image classification. The outcomes of this work are significantly better in terms of accuracy and other factors, which are discussed in section III.

A. Pre-Processing

In X ray mammograms, presence of tags, scratches and pectoral muscles, may interfere in proper extraction of features and result in erroneous classification. So, we have cropped all images depending upon their region of interest (ROI) provided in the database [2]. Further, for smoothing of mammograms Gaussian filter is used. Gaussian Filter is one of the well-known algorithms used in image processing for reducing the additive noise in order to enhance the images.

Gaussian filter uses a 2-D convolution operator which blurs the images by removing details and noise. A Gaussian function is used for transformation of each pixel in the image to a new value.

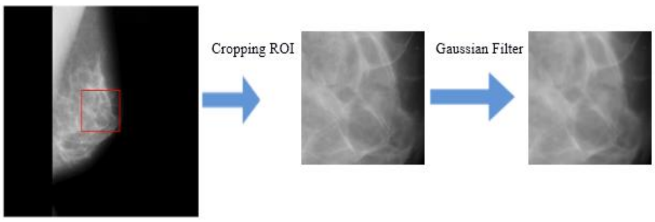


Fig. 1. Pre-processing of mammogram

The equation of Gaussian function in one dimension is as follows:

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \quad (1)$$

In two dimensions, it is the product of two Gaussian functions for x and y respectively:

$$G(x) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (2)$$

where x is the distance of pixel from origin in the horizontal axis, y is the distance of pixel from origin in the vertical axis and σ represents the standard deviation of Gaussian distribution. Figure 1 shows the consecutive outcome of pre-processing steps.

B. Feature Extraction

In this paper, we extract texture features to classify mammograms. One of the ways for representing texture is to use the intensity-histogram based statistical moments of an image [2]. Using only histograms in our analysis for describing texture will capture information only about the distribution of intensities, but not about the spatial relationships between two or more pixels in that texture. Hence in this paper, we use the gray level co-occurrence matrix (GLCM) to extract the second order statistical texture features [4].
GLCM matrix for an image I with n*m dimension, parameterized by an offset $(\Delta x, \Delta y)$ is defined as:

$$C_{\Delta x,\Delta y}(i,j) = \sum_{p=1}^{n} \sum_{q=1}^{m} \begin{cases} 1, \text{if } I(p,q) = i \text{ and } I(p+\Delta x, q+\Delta y) = j \\ 0, \quad \text{otherwise} \end{cases} \quad (3)$$

where i and j are the image intensity values of the image, p and q are the spatial positions in the image I and the offset (Δx, Δy) depends on the direction used θ and the distance d at which the matrix is computed. Changing either of d and θ yields a different co-occurrence matrix. In our approach, we have used matrices with d= [1 to 30] and angles 0, 45, 90 deg., leading to a total of 90 GLCM matrices. From the co-occurrence matrix, we can use several statistics to get more useful set of features. According to Haralick, there are 14 such textual features. However, as proposed by Wei et al. [5] the 4 prevailing features of GLCM on the basis of t-test are Correlation, ASM, Sum Variance and Sum Entropy. Thus, this paper will use focus on these 4 features only. In total, 4*90=360 features are derived from GLCM. This work will use 4 statistical features of GLCM contrast, energy, homogeneity and correlation [2, 4].

Contrast: Contrast indicates change in the intensity contrast of a pixel with its neighboring pixels in an image.

$$Contrast = \sum_{i,j} |i-j|^2 p(i,j) \quad (4)$$

Energy: Energy shows measure of uniformity of texture in an image.

$$Energy = \sum_{i,j} p(i,j)^2 \quad (5)$$

Homogeneity: Homogeneity shows how uniformly the pixels are distributed in an image.

$$Homogeneity = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \quad (6)$$

Correlation: Correlation gives the measure of how each pixel in an image is correlated to its neighbour for entire image

$$Correlation = \sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \sigma_j} \quad (7)$$

where $p(i, j)$ is the element i, j of the normalized symmetrical GLCM and i and j vary from 0 to N-1 where N is the number of gray levels in the image.

C. Feature Selection

In this paper, we use the AdaBoost training process to determine the relative importance of features and use it for performing feature selection. AdaBoost picks best features from data apart from weighting weak classifiers and use them in testing phase to perform classification efficiently. Unlike other feature selection approaches where features are selected by their individual contributions, AdaBoost feature selection method also deals implicitly with the interdependence of features. We, thus, modify the standard AdaBoost algorithm by simultaneously assigning a score to each feature and increasing it based on how many times the corresponding feature gets selected as the optimum classifier at every iteration of the training process. For the feature selection task, we modify the standard AdaBoost algorithm, as proposed by Rojas [11].

Algorithm 1: The simplified AdaBoost algorithm for feature selection

---

Input:

(a) A training image set I with class labels y_I.

(b) The initial feature set, F.

(c) The desired number of iterations, T.

Output:

The importance $<S^{(1)}, \dots, S^{(T)}>$ of features $<F^{(1)}, \dots, F^{(T)}>$.

Proposed Method:

1. F ' := F ;
2. Initialize the initial weights, $w^{[1]} = \{w_I^{[1]}\}_I$, each equal to 1.
3. **for** t = 1 **to** T **do**
4. Normalize the image weights $w^{[t]} = \{w_I^{[t]}\}_I$:

$$w_I^{[t]} := \frac{w_I^{[t]}}{\sum_{J \in I} w_J^{[t]}} \text{ for all } I \in I$$

5. Select feature F^J that minimizes the sum of classification error over all images weighted according to $w^{[t]}$.
6. Increment the value of S^J

$$S^J := S^J + 1$$

7. Update the weights of each image sample I ($w_I^{[t]}$). The image sample weights should be proportional to the error rate E on that image.

The modified AdaBoost procedure, shown in Algorithm 1, iteratively selects a feature as a weak learner initially. In every iteration t, AdaBoost selects the feature (say F^j) that minimizes the sum of training error (achieves the highest classification accuracy) over all images weighted according to w. This feature F^j is selected alone for prediction as a decision stump (or the primary feature in a decision tree).A decision stump is a learning model consisting of a one-level decision tree [10]. The score S^j of the feature F^j is incremented by 1. After that, weight is assigned to each mammogram in the training set proportional to the current error E on the prediction of that mammogram. Hence in this process, the relative influence of images that were correctly classified by the selected feature decreases. The weights of the images misclassified by the weak learner thus increases. These weights can intuitively favour the training of the weak learner, for instance, decision trees can be grown that favour splitting subsets of images with enlarged weights. It also encourages the selection of features that performs well on the misclassified images by the classifier during the previous iteration and is complimentary to the previously selected feature. In this manner, AdaBoost will implicitly deal with feature-correlation in addition to their individual importance. The output of the modified AdaBoost algorithm is the vector S, which is the relative feature importance of the original features. We fix a certain threshold and select all those features whose feature importance is greater than the specified threshold. The value of threshold is varied and accuracy is calculated for each threshold, the results are presented in result section.

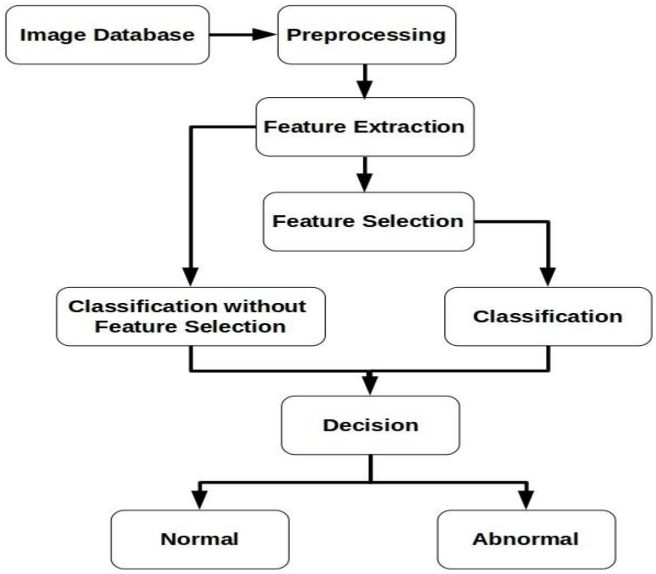


Fig. 2. Proposed working diagram

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

D. Mammogram Classification

The working flowchart of this model is shown in Figure 2, where the database has been pre-processed by cropping the unwanted region of mammograms and suppressing the noise. Texture features of each mammogram are extracted using GLCM. After feature extraction, this paper also examined most relevant features through AdaBoost feature selection method. Further, for the classification of mammogram, relevant and all features are independently tested using random forest classifier. Random Forest, developed by Brieman [13] consists of a set or ensemble of un-pruned decision tree classifiers, with a randomized selection of features at each split. Each decision tree gets a "vote" in predicting the output. In this training process, the training set for each tree is constructed from a bootstrap sample of a training data by sampling N mammogram sample at random, with replacement from the training data. For the purpose of development and assessment of our proposed model, we have trained the classifier multiple times and best accuracy is noted. The classification of the images has been done as into Normal and Abnormal classes.

III. Results Analysis and Discussion

Dataset used for this study is digital mammograms which have been taken from the published Mammography Image Analysis Society (MIAS) database [8, 15]. The original MIAS database was digitalized at 50 micron pixel edge which has been reduced to 200 micron pixel edge which makes every image of 1024 x 1024 pixels. This includes “truth”-markings on those locations where abnormality may be present. Some of the images consist of more than one abnormalities. Therefore, we get a total of 330 images, out of which 207 are normal, 69 are benign and 54 are malignant. So, we have 63% of normal data, 16% of malignant data and 21% of benign data. All images are 8-bit gray level scale images which have 256 different gray levels (0-255) and are in portable gray map (pgm) format. The dataset is divided in 75:25 composition, 75% of data is used for training model and remaining 25% used for testing.

- Metric Used

The performance analysis of this classification framework is done using three statistical parameters, accuracy, specificity and sensitivity. These parameters can be computed using equations (8), (9) and (10) respectively.

$$Accuracy = \frac{TN+TP}{TN+FP+FN+TP} \tag{8}$$

$$Specificity = \frac{TN}{TN+FP} \tag{9}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{10}$$

where TP is true positive, TN is true negative, FP is false positive and FN is false negative. These four values can be obtained from the confusion matrix. From equations (9) and (10), it can be concluded that specificity is the measure of predicting negative test for sample having no disease whereas sensitivity is the measure of predicting positive test for sample having disease. These parameters help in measuring how much each positive or negative predicting model is accurate.

- Discussions

After cropping the images, an analysis is done by finding accuracy score for a set of 4 features derived from GLCM which are contrast, energy, homogeneity and correlation of gray level values for a particular pixel distance *d* and direction θ. As described in Fig. 3, we have used *d*= [1, 30] and angles 0˚, 45˚, 90˚.Thus for each angle θ, we obtain 30 GLCM matrices. The top 4 matrices are selected by individually calculating the miss-classification error in prediction by the feature set of each matrix. Performance of GLCM features depend upon pixel relation between, pixel neighbouring and angles. In Figure 3, we have shown the importance of pixel distance and angle, and select4 best neighbouring pixel distance from 3 angles. After this, we obtain total 12 matrices and their feature set. Each of F_1, F2...F_{12} are GLCM matrix containing four features contrast, energy, homogeneity and correlation. Therefore total features become 3*4*4 = 48 features.

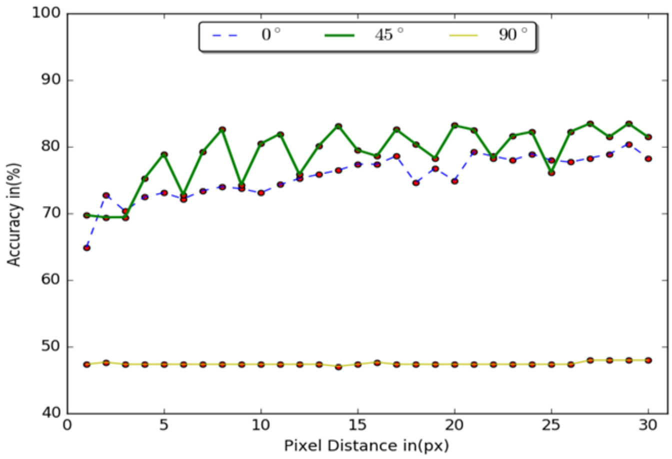


Fig. 3. Performance on different neighboring distance

Table-I Best neighbouring pixel distance and direction

Angle θ	Best four pixel distances
0˚	F1(d=21), F2(d=24), F3(d=28), F4(d=29)
45˚	F5(d=14), F6(d=20), F7(d=27), F8(d=29)
90˚	F9(d=16), F10(d=27), F11(d=28), F12(d=29)

- Mammogram classification without feature selection

For the classification of normal–abnormal mammograms, discrimination analysis of random forest classifier is done using 48 features, extracted from 12 different GLCM matrices as reported in Table I. Number of decision trees are set as 100.

Based on confusion matrix of this classifier, the best achieved accuracy, sensitivity and specificity are 89.02%, 90.56% and specificity 86.20%, respectively.

Table-II Confusion matrix of Normal and Abnormal mammogram classification using RF

		Classification	
		Normal	Abnormal
Truth	Normal	48	5
	Abnormal	4	25

- Mammogram classification with feature selection

AdaBoost feature selection process, as described in above section, is run to select the best scoring features from set of 48 features according to their feature importance. The parameter of AdaBoost Classifier estimators is set to 100.The importance of features can be seen from Figure 4 some of the best scoring features are Energy, Homogeneity features of GLCM matrix of F8 and Correlation of GLCM matrix of F7.

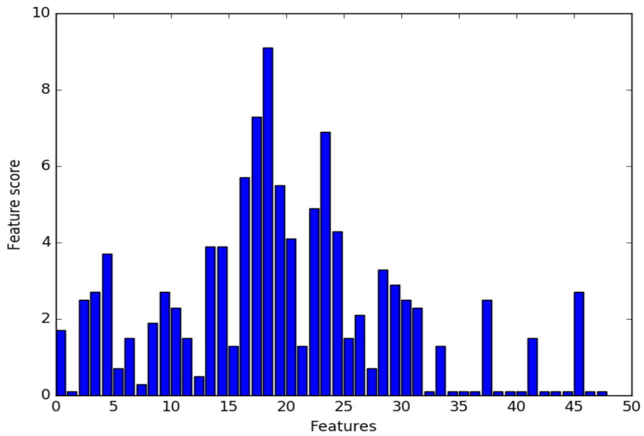


Fig. 4. Feature Importance

We now perform an analysis on accuracy obtained by varying the number of best features selected based on their score as seen in Figure 5. The maximum accuracy of 93.90% is obtained when no. of selected features are 6.These 6 features are Energy, Homogeneity, Contrast, Correlation features of GLCM matrix of F_8 and Correlation, Homogeneity of GLCM matrix of F_7.

Table-III Confusion matrix of Normal and Abnormal mammogram classification using RF with feature selection

		Classification	
		Normal	Abnormal
Truth	Normal	51	2
	Abnormal	3	26

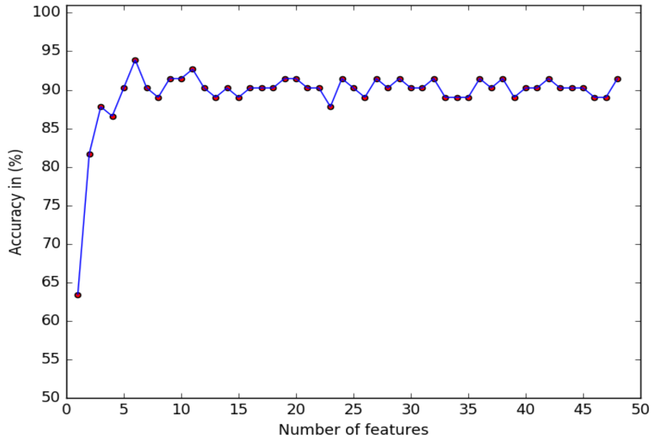


Fig.5. Classifier accuracy

Table-IV Classifier performance on different number of features

Number of Features	Accuracy	Sensitivity	Specificity
2	81.70%	72.41%	86.79%
4	86.58%	79.31%	90.56%
6	93.90%	89.65%	96.22%
8	89.02%	82.75%	92.45%
10	91.46%	86.20%	94.33%
12	90.24%	89.65%	90.56%

Based on these informative features, we trained the random forest classifier using 75 % of database images and test on the remaining 25 % images. Confusion matrix, for test cases are formed and shown in Table-III. Using this confusion matrix, we have calculated the other performance measures. Table-IV, quantify the performance on different number of features. For 6 relevant features, best accuracy, sensitivity and specificity are achieved. The best obtained accuracy; sensitivity and specificity are 93.90%, 89.65%, and 96.22%, respectively. So due to challenging properties of MIAS database where visual appearances of all the mammograms are much closed to each other, classification performance of this work is significantly encouraging.

IV. CONCLUSIONS AND FUTURE WORK

Based on the exhaustive experiments conducted on GLCM matrixes, for finding the best pixel distance and angles. It is concluded that proposed 48 GLCM features based on 3 different angles (0˚, 45˚, and 90˚) from four mentioned pixel distances, classified digital mammograms with 89.02% accuracy, 90.56% sensitivity and 86.40 % specificity using random forest classifier. Further, we examined the 6 most informative features selected by AdaBoost feature selection method, and got significantly encouraging performances,

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

93.90 % accuracy, 96.22% specificity, and 89.65% sensitivity for the normal and cancerous mammograms. In future, we will use this work for the content based retrieval of mammograms, where this work is treated as a pre-processing step for the grouping of the normal and abnormal images; also this model helps in retrieval by detecting the categories of query mammogram.

REFERENCES

[1] Srivastava, S., Sharma, N., Singh, S.K. and Srivastava, R., 2014. Quantitative Analysis of a General Framework of a CAD Tool for Breast Cancer Detection from Mammograms. *Journal of Medical Imaging and Health Informatics*, *4*(5), pp.654-674.

[2] Pratiwi, M., Harefa, J. and Nanda, S., 2015. Mammograms Classification Using Gray-level Co-occurrence Matrix and Radial Basis Function Neural Network. *Procedia Computer Science*, *59*, pp.83-91.

[3] Sahiner, B., Chan, H.P., Petrick, N., Wei, D., Helvie, M.A., Adler, D.D. and Goodsitt, M.M., 1996. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. *Medical Imaging, IEEE Transactions on*, *15*(5), pp.598-610.

[4] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610-621, Nov. 1973.

[5] C.-H. Wei, C.-T. Li, and R. Wilson, "A General Framework for Content-Based Medical Image Retrieval withits Application to Mammograms," in *Proceedings of the SPIE*, vol. 5748, San Diego, CA, 2005, pp. 134-143.

[6] R. Gupta, P.E. Udrill, “The Use of Texture Analysis to Delineate Suspicious Masses in Mammography”, Phys. Med.Biol., VOL.40, PP.835-855 (1995).

[7] Neha Tripathi and S. P. Panda, "A Review on Textural Features Based Computer Aided Diagnostic System for Mammogram Mass Classification Using GLCM & RBFNN," International Journal of Engineering Trends and Technology (IJETT) , vol. 17, no. 9, pp. 462-464, Nov. 2014.

[8] http://www.mammoimage.org/databases/

[9] http://www.breastcancerindia.net/statistics/trends.html

[10] Dietterich, T.G., 2000. Ensemble methods in machine learning. In *Multiple classifier systems* (pp. 1-15). Springer Berlin Heidelberg.

[11] Rojas, R., 2009. AdaBoost and the super bowl of classifiers a tutorial introduction to adaptive boosting. *Freie University, Berlin, Tech. Rep*.

[12] Singh, Vibhav Prakash; Srivastava, Rajeev, Design & performance analysis of content based image retrieval system based on image classification using various feature sets, ABLAZE, Pages: 664 - 670, 2015.

[13] Breiman, L., 2001. Random forests. *Machine learning*, *45*(1), pp.5-32.

[14] Singh, V.P., Gupta, A., Singh, S. and Srivastava, R., 2015, December. An efficient content based image retrieval for normal and abnormal mammograms. In *2015 IEEE UP Section Conference on Electrical Computer and Electronics (UPCON)* (pp. 1-6). IEEE.

[15] Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., et al. (1994). The mammographic image analysis society digital mammogram database. Exerpta Medica, International Congress Series, 1069, 375–378.

[16] Jen, C.C. and Yu, S.S., 2015. Automatic detection of abnormal mammograms in mammographic images. *Expert Systems with Applications*, *42*(6), pp.3048-3055.

[17] Zheng, B., 2009. Computer-aided diagnosis in mammography using content-based image retrieval approaches: current status and future perspectives. *Algorithms*, *2*(2), pp.828-849.

[18] Youlden, D.R., Cramb, S.M., Dunn, N.A., Muller, J.M., Pyke, C.M. and Baade, P.D., 2012. The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality. *Cancer epidemiology*, *36*(3), pp.237-248.

[19] Aarthi, R., Divya, K., Komala, N. and Kavitha, S., 2011, December. Application of Feature Extraction and clustering in mammogram classification using Support Vector Machine. In Advanced Computing (ICoAC), 2011 Third International Conference on (pp. 62-67). IEEE.

Enhancement of Intrusion-Detection System in MANETs with the Digital Signature via Elliptic Curve Cryptosystem

K. Spurthy
Computer Science & Engineering
KL University, AP, INDIA

T. N. Shankar
Computer Science & Engineering
KL University, AP, INDIA

S.Sabari Giri Murugan
Computer Science & Engineering
KL University, AP.INDIA

Abstract- **The watchdog scheme is popular in MANET to defend the malicious attacks, but the major pitfall of this method is unable to detect some destructive actions. The technique Enhanced adaptive acknowledgment EAACK is designed to handle some weaknesses as false misbehavior, limited transmission power, and receiver collision of the watchdog scheme that is not fully efficient to resolve all the problems. This paper focuses intrusion detection system on MANETs with the collaboration of three IDS approach and with the techniques ACK, 2-ACK, and misbehavior report identification MRI. This paper proposes digital signature with Elliptic Curve Cryptosystem to avoid forging acknowledgment packets from attackers.**

Keywords: **DSR, MANET, AOMDV, watchdog, ACK, 2-ACK, MRI.**

1 INTRODUCTION

The security system in Mobile adhoc network (MANET) is one of the emerging topics as a research area of networking. The collection of mobile nodes is known as MANET which communicates via wireless radio links and nodes without any central base station. The responsibility of each node in the network is to forward data packets for other nodes. A node may leave a network or rejoin, and free to move in any direction, because of the dynamic topology. Packet routing towards the adjacent nodes is one of the tough tasks in a network for regular changes in topology. The vulnerability of the malicious attacks is more to the MANETs for its flexibility and adaptability nature. Entire routing protocol should be designed and chosen in such a way to provide high reliability, security, power efficient, avoid overhead and best quality of service [1, 3]. Authentication and encryption must be used as the primary defense to protect the network. The security is a very much essential with a good impact on the performance of any network. Generally Intrusion Detection System (or IDS) has been used to detect any malicious threat to the systems.

Less security in a network is a major cause of interruption in transmission which leads to more energy consumption and hazardous on the data transmission between the mobile nodes. The attacks which are specific to the data transmission process and its defense via elliptic curve cryptosystem (ECC) to defend an intruder is one of the emerging topics in networking focused in this paper. One of the main attacks against ad hoc networks affecting their routing protocols are named routing-disruption attacks can be overcome by the ECC authentication system. The elliptic curve digital signature scheme algorithm(ECDSA) in Adhoc on-demand Multipath Distance Vector (AOMDV) to detect the forge acknowledgement packets is proposed in this paper which leads to a better throughput with less routing overhead.

2 BACK GROUND

Elhadi has discussed on intrusion detection to overcome the Intruder attacks. Later Siddiqi and Hymavathi have stressed on secure intrusion detection system for MANETS by using Enhanced adaptive acknowledgement EAACK technique as a novel contribution [1,3].The defence technique to thuwrt the Intruders via watchdog approach is published by Nidhi Lal [2]. AOMDV which is an extension to AODV protocol for computing multiple loop-free and link disjoint paths is discussed by Mahesh and Samir [4].

A simulation and comparison of AODV, DSR and AOMDV routing protocols to prove the fastest process of detection path are present in the paper by Chaddha et al. [5]. ECDSA has a vital role to detect the intrusion via digital signature of images are demonstrated by Shankar et al. [6,7]. The way to provide an improved security system to the MANET is highlighted by Perrig and Johnson [8]. This paper is developed after a comprehensive study work of these referred articles.

3.1 Adhoc on-demand Multipath Distance Vector (AOMDV)

In Ad hoc on demand distance vector AODV routing protocol is necessary to decrease high overhead and latency at mobile nodes. "Ad hoc on demand distance multiple distance vector" AOMDV is introduced to resolve this issue which is available with the multiple paths.

AOMDV protocol is the extension of AODV specially designed to calculate multiple paths at the time of route discovery. This process is highly dynamic which causes frequent link failure among the moving nodes. Route discovery procedure is one of the major issues in this process and it happens in case of overall paths to source or destination fails.

An alternate path to the source or destination is defined by route procedure. Such types of copies may be a chance to form routing loops. A similar invariant should be maintained as it defines in single path case to overcome the situations. However, the major inequality in the multiple next-hop routes obtained by multiple route advertisement are accepted and maintained as long as the invariant is to be complied. Various routes to the same destination may have different hop-counts is a probable drawback. Therefore, route identification is essential to determine which of these hop-counts is advertised to other or the unfeasibility of advertising, different hop-count to different neighbors with the same destination sequence number.
The advertised hop-count is defined as the maximum hop-count of the multiple paths for destination available at source cannot be changed against the maximum hop-count for the same sequence number. AOMDV follows such type of techniques to find multiple paths.

AODV	AOMDV
Destination	Destination
Sequence number	Sequence number
Hop-count	Advertised Hop-count
Next Hop	Route-list{(nexthop1,hop-count1), (nexthop2,hop-ount2),…}
Expiration-timeout	Expiration-timeout

Table 1 AODV Routing

Table 1 describes the structure of routing entries for AODV and AOMDV.

3.2 Dynamic Source Routing (DSR)

Dynamic Source Routing (DSR) is a routing protocol, which is similar to AODV in that it forms a route on-demand when a transmitting computer requests one.

DSR protocol uses source routing instead of using routing table at each intermediate device, and the routing information is maintained (continually updated) at mobile nodes. Determining source routes requires each node appends own identifier when forwarding RREQ during route discovery. The appended path information is caught by nodes processing the route discovery packets. The routed packet contains the address of each device to minimize the overhead cost to traverse a long distance. DSR optionally defines a flow id option that allows packets to forward on a hop-by-hop basis to avoid the source routing.

Route discovery and route maintenance are the two major phases of DSR protocol. Route reply would only be generated as soon as the message reaching at destination node.
The destination node should return the route reply. The route would be used if the route is in the destination node's route cache, else the node reverse the route based on the route record in the route reply message header.

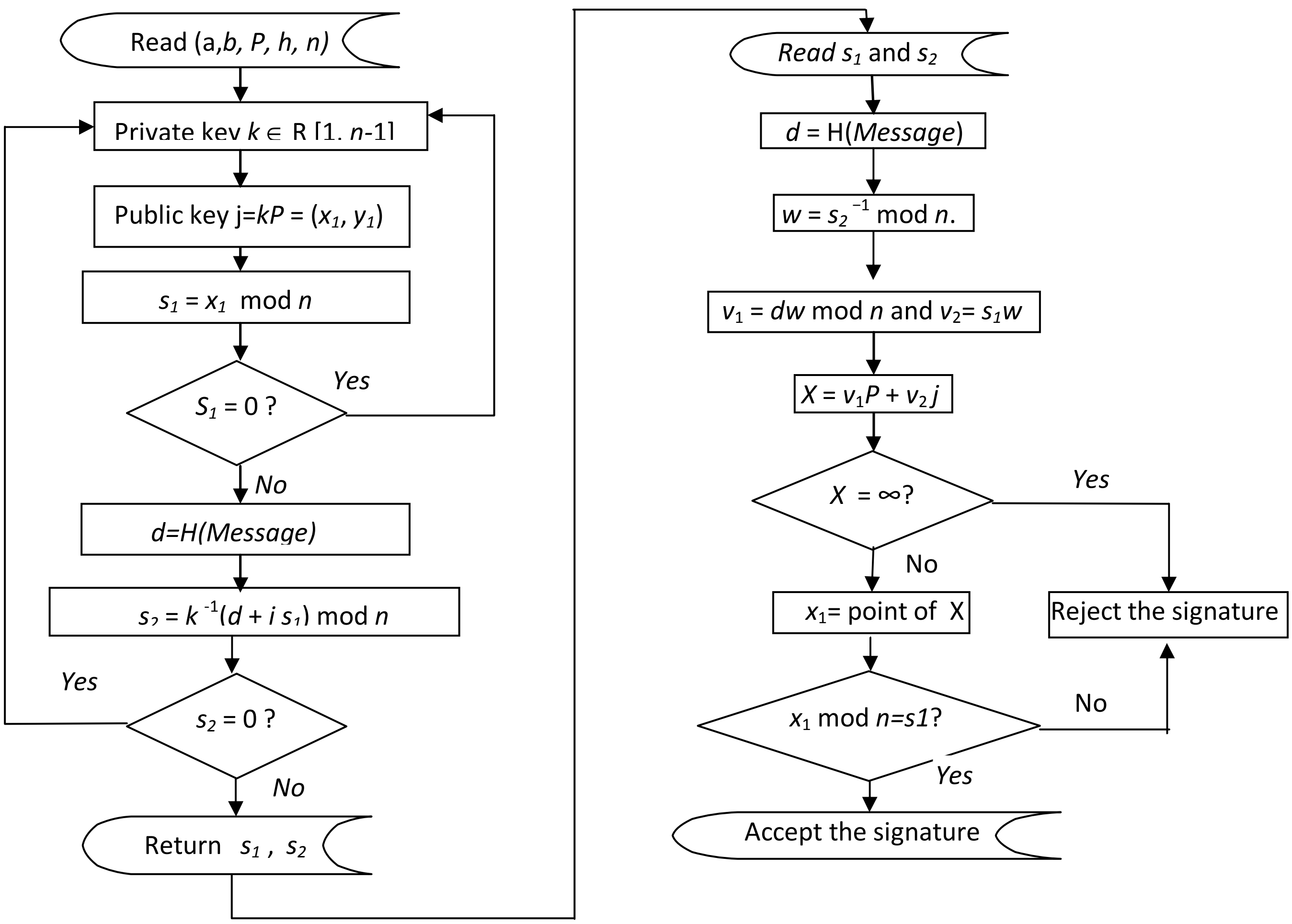
The route maintenance phase is initiated with the route error packets are generated at a node. The error generated hop should be removed from the node's route cache and all routes containing the hop are truncated at that point. Again, the route discovery phase should be initiate.

4 Digital Signature

Digital signature is a mechanism used to protect the information. Digital signature can be prepared by using RSA, Elagamal or Elliptic curve crypto system (ECC). Among those Elliptic Curve Digital Signature Algorithm (ECDSA) is preferred for limited working space with the nodes. It can provide higher level security with the fast processing speed. A comparison with Elgamal digital signature is present in this paper to prove the speed and compatibility of ECDSA.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

4.1 Elliptic Curve Digital Signature Algorithm



S-node *D*-node

Fig.1. Communication with digital signature.

This paper focuses on digital signature via ECC. The main purpose of this implementation is to compare the performance of both algorithms in MANETs with the digital signature via ECC and Elgamal. For every message m, a pre agreed hash function H is applied. According to ECDSA algorithm it has to read a,b,h,n. Then assume the private key k belongs to 0 to n-1. Multiply k with a point P which is lying on the Elliptic curve to generate the public key.

Assume $s_1 = x_1$ and s_2 can be computed as $s_2 = k^{-1}(d + i\ s_1)$ mod n. s_1 and s_2 have to send the destination after embedding with ACK. ACK should be verified node to node but digital signature should be verified at the destination node to prove the malicious attack. Send s_1 and s_2 with the message to the destination node for signature verification. Entire message has to pass through the hash algorithm to obtain the message digest d. Assume $w = s_2^{-1}$ mod n and calculate $v_1 = dw$ mod n and $v_2 = s_1 w$. Compute $X = v_1 P + v_2 j$. Assume x_1 as value of X coordinate. If x_1 mod $n = s_1$ accept the signature else reject it.

4.2 Elgamal Digital signature scheme

According to the signature scheme choose the private key x and calculate the public key as g^x mod p. For every message m, pick a random number k and compute $S_1=g^k$ mod p and $S_2=k^{-1}(m\text{-}xS_1)$ mod p. Now send the signature (S_1, S_2) to the receiver by embedding with the message m. Verify the signature as $g^m=P_k^{S1}S1^{S2}$**IDS Systems**: The researches had provided number of collaborative IDS systems namely

1. Watch dog
2. ACK
3. 2-ACK

4.3 Watchdog

Watchdog IDS works by listening the transmission to its next hop to detect malicious behaviour. It increases its failure counter with overhear of the next node fails to forward the packet within a certain period of time. The watchdog node reports a misbehaving when failure counter of a node exceeds a predefined threshold.

5 PROBLEM DEFINITION

Three of the six weaknesses of Watchdog scheme: False misbehaviour, limited transmission power, and receiver collision are handled by the proposed system. This section is stressed on these three weaknesses in detail.

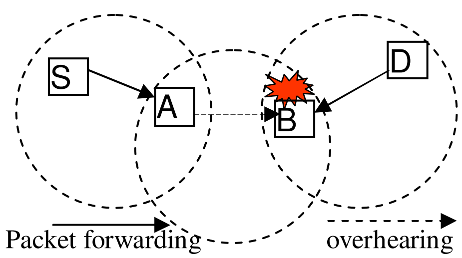


Fig 2 Receiver collision

In Figure 2 both nodes A and D are trying to send packets to node B at the same time node B is unable to receive the both the data.

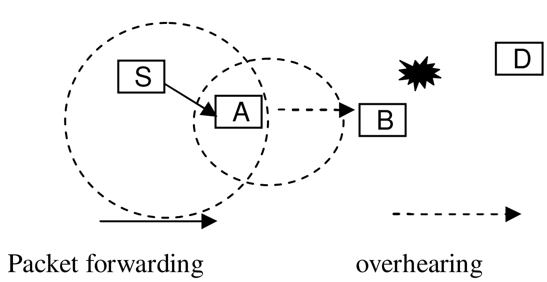


Fig 3 Limited transmission power:

In Figure 3 node A limits its transmission power so that the packet transmission can be overheard by node A, but it too weak to reach node B.

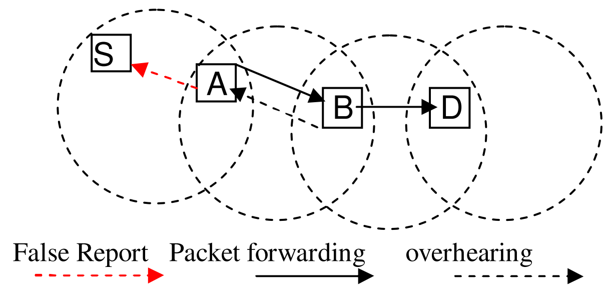


Fig 4 False misbehaviour report

In Figure 4 Node A sends back a misbehaviour report even though node B forwarded the packet to node D.
2-ACK and ACK solves receiver collision and limited transmission power was discussed. However, these two are vulnerable to the false misbehaviour attack. This technique adopts the digital signature scheme during the packet transmission process to provide more security and overcome the problem of false misbehaviour within the process finding acknowledgment-based IDSs.

1 ACK

The ACK is an end-to-end acknowledgment scheme. In the ACK scheme, the intermediate nodes simply forward the packet1 which sends out by the source node S. As soon as destination node D receives the Packet 1, it requires sending back an ACK acknowledgment packet to the source node S in a reverse order of the same route within a predefined period of time. Otherwise it must be switched to 2-ACK scheme.

2 2-ACK

Misbehaving links can be identified by acknowledging every data packet transmitted over every three consecutive nodes along with the path from the source to the destination. First of all Node A forwards Packet 1 to node B, then node B forwards Packet 1 to node C. Now Node C is two hops away from node A and it is ready to generate a 2-ACK packet after receiving the Packet1. Whenever the node C sends a 2-ACK packet back to A then it is a successful transmission of Packet 1 from node A to node C. If 2-ACK packet is not received in a predefined time period, both node B and node C are reported as malicious. The same process applies to every three consecutive nodes along with the rest of the route.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

The misbehaviour report is generated by node A and sends to the source node S. This technique requires the source node to switch in MRI mode to confirm this misbehaviour report. This is a critical step to detect false misbehaviour report.

3 MRI

The misbehaviour report identification (MRI) is designed to identify the misbehaviour report in the presence of false misbehaviour report. This resolves the weakness of Watchdog when it fails to detect misbehaving nodes. In general the false report can be generated by malicious attackers to report the innocent nodes as malicious nodes. This type of attacks will cause great harm or destruct the network. The attack is dangerous when the attackers break down number of nodes that cause a network division. The main aim of MRI scheme is to authenticate whenever the destination node is received the missing packet through different paths. To initiate the MRI mode, the source node first searches an alternative route to the destination node by using its local knowledge. If there is no alternative path then, the source node starts an AOMDV routing request to find another route. The misbehaviour reporter node can be stopped by adopting an alternative route to the destination node. The destination node searches its local knowledge base and compares if the reported packet was received or not as soon as receiving an MRI packet. Whoever generated this report is marked as malicious, if it is already received after that safe to conclude as a false misbehaviour report. Or else, the misbehaviour report is trusted and accepted.

6 SCHEME DESCRIPTION

Intrusion detection system on mobile ad hoc networks (MANETs) with the collaboration of three intrusion detection system (IDS): ACK, secure 2-ACK, and misbehaviour report identification (MRI) are focused in this work. Digital signature is introduced to avoid forging acknowledgment packets from attackers.

In proposed system a comparison of AOMDV and DSR protocol is present with performance metrics like packet delivery ratio and routing overhead. In a network the source and destination nodes are assumed as trusted nodes, and they act as both sender and receiver.

All acknowledgment packets in a network system should be digitally signed at source node and verified by destination node. To distinguish different packet types(ACK,2-ACK and MRI) 2-bit packet header should be added.

Packet Type	Packet Flag
General Data	00
ACK	01
2-ACK	10
MRI	11

Table 1 Packet Type Indicators

7 PERFORMANCE EVALUATION

This section is focused on, simulation environment and methodology as well as comparing performances through simulation result comparison with AOMDV and DSR.

7.1 Simulation Methodologies

To measure the performance of proposed system under different types of attacks, three states have been proposed to simulate different types of misbehaviours or attacks.

State 1: In this state, a basic packet dropping attack is simulated. Malicious nodes simply drop all the packets that they receive. The purpose of this state is to test the performance of IDSs against two weaknesses of Watchdog, namely, receiver collision and limited transmission power.

State 2: This state is exclusively designed to test IDSs' performances against false misbehaviour report. Malicious nodes always drop the packets that they receive and send back a false misbehaviour report whenever it is possible in case of it.

State 3: This state is intended to test the IDSs' performances when the attackers are smart enough to tamper acknowledgment packets to claim positive result while, actually, it is negative. As Watchdog is not an acknowledgment-based scheme, it is not eligible for this state setting.

Algorithm:

Step 1: Start acknowledgement mode
Step 2: Check the node activity in packet mode
Step 3: If packet mode is ACK, then check whether reply is from destination or not.
Step 4: Else switch to 2-ACK mode.
Step 5: Go to step 2.
Step 6: If the packet mode is 2-ACK then check node misbehaviour if yes send the MRI
Step 7: Else send ACK.
Step 8: Go to step 2.
Step 9: If the packet mode is MRI then check the content of destination node.

Step 10: If data existed then mark reported node is malicious.
Step 11. Else send ACK.
Step 12. Go back to *Step* 2.

7.2 Simulation Configurations

The simulation is conducted within the Network Simulator (NS) 2.34 environment on a platform. The system is running on a system with Core i3 and 4-GB RAM.

S.NO	Parameter	Value
1	Simulator	NS-2
2	Channel type	Channel/Wireless Channel
3	Radio Propagation Model	Propagation/Two Ray Ground
4	Network interface Type	Phy/WirelessPhy
5	MAC Type	Mac /802_11
6	Interface Queue Type	Queue/Drop Tail/ PriQueue
7	Routing Protocol	DSR,AOMDV
8	Antenna	Antenna / Omni Antenna
9	Type of traffic	CBR
10	Area (M*M)	1216 *743
11	Simulation Time	50 sec
12	No of nodes	18

Table 2 Simulation Parameters

1) Packet delivery ratio (PDR): PDR is the ratio of the number of delivered packets to the destinations divided by the total number of packets actually sent. The greater the value of the packet delivery ratio, the better is the performance of the protocol.

2) Routing overhead (RO): The additional costs incurred during the data packet delivery process. It contains routing-related transmissions [Route request (RREQ), Route reply (RREP), Route error (RERR), ACK, 2-ACK, and MRI].

3) Throughput (t): throughput is defined as the rate of successful message delivery over a communication channel.

7.3 Performances comparison between ECDSA and Elgamal digital signature(EDS) schemes

The Elagamal DSA key length is 1024bits in other hand ECDSA key size is 120 bits. ECDSA is faster than EDS for its smaller key size, which is suitable to the nodes of a MANET. The run time is less with better security to transfer the packets hop by hop over a network.

7.4 Observations

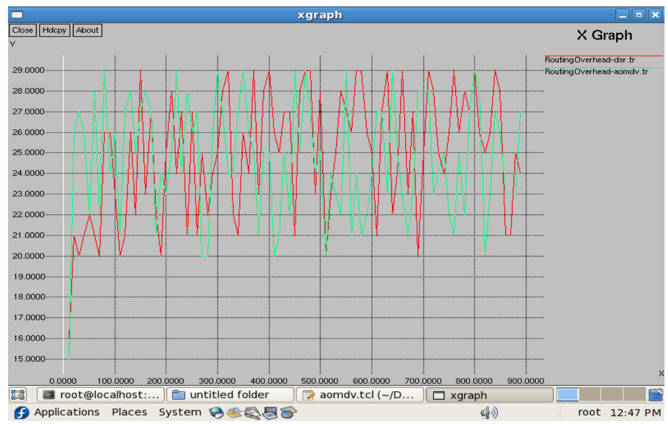


Fig 5 comparing over head of AOMDV and DSR

In Figure 5 red color indicates the routing overhead in AOMDV routing protocol and the green color indicates the routing overhead in DSR routing protocol. when node moving frequently the routing overhead in AOMDV performance is less compare to DSR

Results	AOMDV	DSR
Number of packets sent	14450	6153
Number of packets received	6772	4927
Throughput	213.379	124.88
Total energy required for nodes communicating	791	1695
Average energy required by each node	7.98	17.12

Table 3 Comparison of Throughput and Energy

From Table 3 it concludes that the throughput is high in proactive routing protocol compared to the reactive routing protocol and nodes energy consumption is less in AOMDV than DSR.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

8 CONCLUSION AND FUTURE WORK

In this paper the AOMDV protocol with digital signature via ECC has been proposed to achieve the enhanced security. Of course the overall performance of AOMDV is better than DSR, but performance cannot be degraded by using ECDSA. It is clear that reactive routing protocols are less preferable than proactive routing protocols in terms of overhead cost and speed. Only fixed numbers of nodes have been considered yet, no emphasis on mobility with neglected pause time. Find out the factors which are responsible for these simulation results, as well as better performance of AOMDV in various situations as compared to DSR are under the development. Further simulation needs to be carried out for the performance evaluation with not only increased number of nodes but also varying other related parameters like Pause Time, Network load, Speed.

ACKNOWLEDGEMENT

This work is supported by the Department of Science and Technology, India through the fund sanctioned for improvement of Science & Technology infrastructure, at department of CSE, K.L University, by order number SR/FST/ESI-332/2013.

REFERENCES

[1] EACK—A Secure Intrusion-Detection System for MANETs Elhadi M. Shakshuki, Senior Member, IEEE, Nan Kang, and Tarek R. Sheltami, Member. IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, VOL. 60, NO. 3, MARCH 2013.

[2] Nidhi Lal "An Effective Approach for Mobile ad hoc Network via I-Watchdog Protocol", (IJETCSE) Volume 10 Issue 1 –OCTOBER 2014.

[3] H. Syed Siddiq1, M. Hymavathi, "EAACK – To Overcome from Intruders Attacks in Manet's by Providing Security Checks", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Impact Factor (2012): 3.358.

[4] Mahesh K. Marina1*,† and Samir R. Das2 "Ad hoc on-demand multipath distance vector routing", Published online in Wiley InterScience (www.interscience.wiley.com).

[5] Manveen Singh Chadha, Rambir Joon, Sandeep "Simulation and Comparison of AODV, DSR and AOMDV Routing Protocols in MANETs", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, July 2012.

[6] T. N. Shankar, G. Sahoo, S. Niranjan, 'Using the Digital Signature of a Fingerprint by an Elliptic Curve Cryptosystem for Enhanced Authentication' , Information Security Journal , A Global Perspective, Taylor and Francis, Vol.21, No. 5, pp. 242-254.

[7] K. Spurthy, T.N.Shankar, Ranjan K. Senapati, 'Improving Authentication of an Iris Recognition System by Digital Signature via Elliptic Curve Cryptosystem', International Journal of Computer Science and Information Security (IJCSIS), Vol. 14, No. 5, pp.135-141.

[8] Y. Hu, A. Perrig, and D. Johnson, —ARIADNE: A secure on-demand rout-ing protocol for ad hoc networks,‖ in Proc. 8th ACM Int. Conf. MobiCom, Atlanta, GA, 2002, pp. 12–23.

K. Spurthy holds B.Tech(CSE) from Adams Engineering College, Khammam and M.Tech(CSE) at Swarna Bharathi Institute of Science and Technology, Khammam, AP, India. She is pursuing Ph.D(CSE) at KL University, Guntur Dist, Andhra Pradesh, India. Her research interest is Intrusion detection in Biometrics.

T.N.Shankar has obtained his M.Tech and Ph.D computer Science & Engineering from Birla Institute of Technology, Mesra, Ranchi, India. At present he is working as Professor, Computer Science & Engineering at KL University, Guntur, AP, India. He has about twenty five publications in his credit. His research interests include Information Security and Soft Computing. He has published a book title "Neural Networks", UNIVERSITY SCIENCE PRESS, New Delhi. Currently eight research scholars are pursuing Ph.D under his guidance. He is a life member of various professional societies such as ISTE and ACM.

S.Sabari Giri Murugan holds B.Tech (IT) from PSNA College of Engineering and Technology, Dindigul and M.Tech (IT) at Sathyabama University, Chennai, India. At present he is working as Assistant Professor, Computer Science & Engineering at KL University, Guntur, AP, India, His research interests include Information Security and Network Security.

P-Method: Improving AODV Routing Protocol for Against Network Layer Attacks in Mobile Ad-Hoc Networks

Shahram zandiyan and Reza fotohi
Department of Computer Engineering
Germi branch, Islamic Azad university
Germi, Iran

Marzieh koravand
Department of Computer Engineering
Germi branch, Islamic Azad university
Germi, Iran

Abstract— **Mobile ad hoc networks are regarded as a group of networks consisted of wireless systems which developing together a network with self-arrangement capability. no constant communication infrastructure and use central nodes to communicate with other nodes. Despite lots of advantages, these networks face severe security challenges, since their channels are wireless and each node is connected to central node. One of these concerns is the incidence of network layer attacks (Black and worm hole attack) is one kind of routing disturbing attacks and can bring great damage to the network. In this attack, an attacker cheats nodes, absorbs their packets and then deletes them. Hence, black hole and wormhole disrupts communication, or even makes it impossible in some cases. In this paper, we proposed P-Method for against network layer attacks in mobile Ad-Hoc networks based on hop count and RTT test. The proposed algorithm is implemented in ns2.35 environments and is compared with AODV And DSR under attacks, and improved AODV in different scenarios. Simulation results revealed that the (P-method), is better than AODV And DSR under attack in terms of packet dropped, packet loss, throughput, and jitter.**

Keywords- Mobile ad hoc networks, AODV and DSR routing protocol, Black hole attack, Worm hole, P-Method.

I. INTRODUCTION

Mobile Ad-hoc network is one kind of new wireless network structures. Unlike devices in traditional Wireless LAN solutions, all nodes are movable and the topology of the network is changing dynamically. Similar to other systems, there is a risk of external agent infiltration in the mobile ad hoc networks. In these networks there is basically no infrastructure, that is no routing devices such as router or switch is used. So, they are highly posed to the risk of various.one the most common attacks in MANET is Black hole attack and worm hole; black hole and worm hole attack, a malicious node uses its routing protocol in order to with the release of false news, having the shortest path to the destination node or to the packet it wants to avoid. This black hole node advertises its Availability of fresh routes irrespective of checking its routing table. In the attacker node will always have the availability in replying to the route request and thus intercept the data packet and retain it [1]. Under such an attack, by misusing routing algorithm packets, attacker node absorbs network traffic and upon receiving packets, instead of forwarding, discards them silently K. [2]. Black hole attack is investigated, and evaluated in different previous works. However, what is simulated or implemented so far does not represent its most negative effect on dropping data packets. In most of related works [3], [4] Absorbing network traffic through malicious node, has been performed using false RREPs in response to received RREQs. The mobile ad hoc networks are not having the fixed network topology due to the reason that mobile nodes are frequently changing their positions and movement. Network topology for the MANET networks is not fixed because of the frequent nodes movement in the network. Mobile ad hoc networks having different types of routing protocols like reactive, hybrid, and proactive protocols type of routing protocols. We can use these protocols with different network scenarios and mobility patterns. The reactive protocols such as DSR4 protocol and AODV5, protocol are frequently used MANET protocols [5 and 29]. In this paper, we choose DSR and AODV as a sample example, because it is one of the protocols being considered for standardization for mobile ad hoc networks. There are other routing protocols, and there are parts of mobile ad hoc networks other than routing that need detection black hole attacks, for example medium Achieve control protocols. We believe the main elements of our method would also apply there, the selfish or black hole and worm hole behavior of nodes can affect routing and network performance. This paper assumes that an efficient defense can

be based on hop count and RTT We implement the proposed scheme over AODV routing protocol in ns-2 2.35 environment. Organized as follows. Section 2 Related work. Section 3 describes the AODV routing protocol. In section 4 describes the DSR routing protocol. In section 5. Attacks in Mobil ad hoc network. Section 6. The Proposed Method against the attacks of the black hole and wormhole attack. Section 7 Compare the proposed method with DSR and AODV under attack Section 8 Experimental Data and analysis. Finally, section 9 concluding.

II. Related work

MANET is very much popular due to the fact that these networks are dynamic, infrastructure less and scalable. Despite the fact of popularity of MANET, these networks are very much exposed to attacks [6]. Wireless links also makes the MANET more susceptible to attacks which make it easier for the attacker to go inside the network and get access to the ongoing communication [7]. Different kinds of attacks have been analyzed in MANET and their effect on the network. Attack such as gray hole, worm hole, black hole, where the attacker node behaves maliciously for the time until the packets are dropped and then switch to their normal behavior [8]. MANETs routing protocols are also being exploited by the attackers in the form of flooding attack, which is done by the attacker either by using RREQ or data flooding [9]. Design and presentation of different security obstacles and attacks in mobile ad hoc networks as well as finding appropriate solutions against them is a challenging research area for researchers. Black hole attack is one of the famous related attacks. In [10], black hole attack is evaluated in DSR based networks and a solution is proposed to mitigate it, as well. In such papers, fake routes are only suggested in response to RREQ packets. In [11], and [12] Black hole attack is assessed in DSR based networks and in [13], is considered in AODV based networks. In such works, fake RREP suggestions are just offered based upon received RREQs, too. Black hole attack operates in [14] in two different phases. It works by both propagating fake RREQs, and generating RREQ based false RREPs.In black hole attack, a malicious node uses its routing protocol in order to advertise itself for having the shortest path to the destination node or to the packet it wants to separate. This malicious node advertises its availability of fresh routes irrespective of checking its routing table. In this way attacker node will always have the availability in replying to the route request and thus intercept the data packet and retain it [15, 16]. Using AODV [17] protocol. But they do not consider the data packets. Instead they consider only control packets like RREQ, RRER and RREP and network layer acknowledgement. A black hole can even drop data packets by perfectly transmitting control packets. There the system fails by thinking there is no black hole as the control packets are transmitted without any delay or drop. Many approaches to detect the black hole attack and to defend The MANET from the attack have been proposed [18]-[19]. According to the algorithm by Deng et al. [20], every node crosses check with its next hop node on the route to the destination on receiving or overhearing a RREP packet. If the next hop node does not have a link to the node that sent the RREP, then the node that sent the RREP is considered as malicious. This solution assumes that there exists at most one malicious node and thus cannot cover the case with two or more malicious nodes, which is quite possible in real situations. An algorithm presented in [21] claims to detect the black hole attack in a MANET which is based on relationships of a cream trust level among the nodes. However, in the real network, it is very difficult to set an appropriate value for the trust level. In the method [22], every node has a function of learning the traffic flow in the network and evaluating the possibility criterion of black hole attack based on such learning results in order to detect the malicious node. If the value of the criterion is larger than a predetermined threshold, the node judges that there exists a black hole attacker. This method only provides detection of a single black hole attacker and cannot detect a chain of malicious nodes which cooperate with each other. TAODV [23] every node has a function of learning the traffic flow in the network and evaluating the possibility criterion of black hole attack based on such learning results in order to detect the malicious node. If the value of the criterion is larger than a predetermined threshold, the node judges that there exists a black hole attacker. This method only provides detection of a single black hole attacker and cannot detect a chain of malicious nodes which cooperate with each other. It calculated the average sequence number and try to find out the malicious node, as the malicious node will send RREP messages with extremely higher sequence number. There are chances of getting RREP packet with highest sequence number from a genuine node too. trust value on the neighbor is being increased or decreased dynamically. The method is implemented only on AODV protocol

III. DESCRIBES THE AODV ROUTING

The Ad Hoc On-Demand Distance Vector (AODV) routing protocol is an adaptation of the DSDV protocol for dynamic link conditions [24] [25]. Every node in an Ad-hoc network maintains a routing table, which contains information about the route to a particular destination. Whenever a packet is to be sent by a node, it first checks with its routing table to determine whether a route to the destination is already available. If so, it uses that route to send the packets to the destination. If a route is not available or the previously entered route is inactivated, then the node initiates a route discovery process. A RREQ (Route Request) packet is broadcasted by the node. Every node that receives the RREQ packet first checks if it is the destination for that packet and if so, it sends back an RREP (Route Reply) packet. If it is not the destination, then it checks with its routing table to determine if it has got a route to the destination. If not, it relays the RREQ packet by broadcasting it to its neighbors. If its routing table does contain an entry to the destination, then the next step is the comparison of the 'Destination Sequence' number in its routing table to that present in the RREQ packet. This Destination Sequence number is the sequence number of the

ISSN 1947-5500

last sent packet from the destination to the source. If the destination sequence number present in the routing table is lesser than or equal to the one contained in the RREQ packet, then the node relays the request further to its neighbors. If the number in the routing table is higher than the number in the packet, it denotes that the route is a 'fresh route' and packets can be sent through this route. This intermediate node then sends a RREP packet to the node through which it received the RREQ packet. The RREP packet gets relayed back to the source through the reverse route. The source node then updates its routing table and sends its packet through this route. During the operation, if any node identifies a link failure it sends a RERR (Route Error) packet to all other nodes that uses this link for their communication to other nodes. This is illustrated in figure 1 and 2.

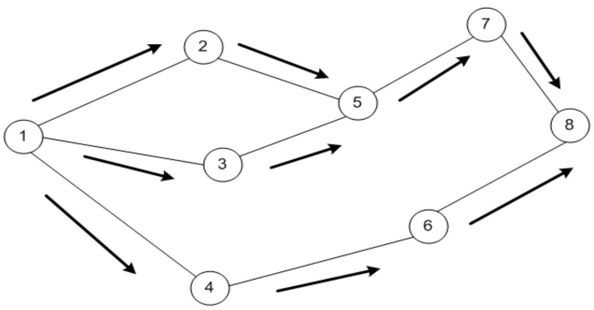


Fig. 1.Broadcast to AODV route discovery

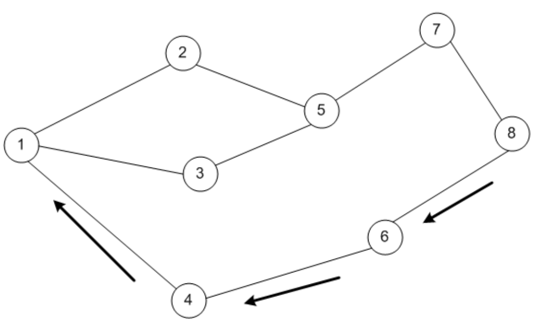


Fig. 2.A sample of route discovery in AODV protocol

IV. DESCRIBES THE DSR ROUTING

DSR protocol is a reactive routing algorithm designed for mobile adhoc network. The process of routing in DSR is composed of two main phases known as route discovery and route maintenance. Routing in DSR is completely carried out in an on-demand method. Route discovery phase is a process under which source node, in order to send data packets, obtains a valid route to the destination node. For this, source node creates a RREQ packet and relays it in the network. Such a packet will be received by all of the sources neighbor nodes. Each RREQ packet contains an identifier and a list of addresses of intermediate nodes which this packet has passed from them. Such a list is initially empty at the time of creating RREQ by the source node. When a node receives a RREQ packet, creates a RREP regarding information included in the list of addresses inside the packet and sends it back to the source node if only it be the destination node itself or have had a route to the destination. Once source node receives such a RREP packet, it first adds this route to its route cache and then starts to send data packets using the route included in the packet. If the receiver of RREQ has not had a route the destination and has not previously received this RREQ packet, appends its address to the list of nodes inside the packet and rebroadcasts that. When the destination node receives a RREQ, it can create and send back the RREP to the source node using the route which can be computed by inversing the list of addresses inside the RREQ packet. Route maintenance is a mechanism by which, as source node is using a route to send its data packets, can discover changes of topology and send remainder of its packets through an alternative route if it be convinced that the current route has been broken and not usable anymore.

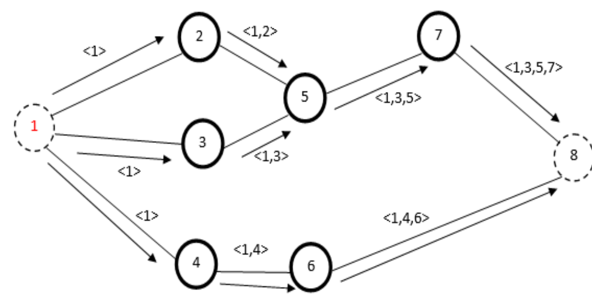


Fig. 3. depicts a discovery route in DSR protocol. (All-over distribution)

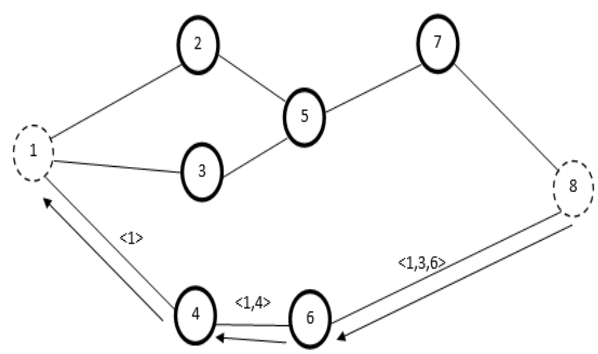


Fig. 4. A sample of route discovery in DSR protocol

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

V. TWO TYPES OF ATTACK IN MOBILE AD HOCE NETWORK

Attacks in MANET can be divided into two classes, according to the Criteria that whether the disrupt the operation of a routing protocol or not. These two classes are passive attacks and active attacks. In a inactive attack, the operation of the routing protocol is not disrupted by the attacker, and only attempts to discover valuable information by listening to the routing traffic is being done. Active attacks, however, involve actions like modification and deletion of exchanging data to absorb packets destined to other nodes to the attacker for analyzing or disabling the network. Some typical kinds of active attacks can be easily performed against MANET, regarded as, flooding attack, selfing attack, gray hole attack, rushing attck, spoofing, wormhole attack, sleep deprivation and impersonation (Jathe et al, 2012) As mentioned, weak infrastructure in mobile ad-hoc networks exposed them to a large amount of attacks. One of these attacks is the black hole attack (Dokure et al, 2006) Black Hole in the (network layer attacks): all packets dropped by a Forged routing packets, the attacker can route all packets for some destination themselves and then discard them

A. Worm hole attack

Wormhole attack which is considered as a severe attack in mobile ad hoc network. Minimum two malicious nodes are required to perform this attack; more than two malicious nodes are also used to perform this attack. In this attack the two malicious nodes reside in the two ends of the network and they form a link between them using an out-of-band hidden channel like wired link, packet encapsulation or high power radio transmission range [26]. After they form a tunnel between them as shown in figure 1, whenever a malicious node receives packets it tunnels them to the other malicious node and in turn it broadcasts the packet there. Since the packet is travelling through the tunnel it reaches the destination speeder than other route and moreover the hop count through this path is going to be less so this path is established between the source and the destination [27]. Once the path is established between the source and the destination through wormhole link they can misbehave in many ways in the network like continuously dropping the packets, selective dropping the packets, analyzing the traffic and performing Denial of Service attack. Figure 5 shows an example of this attack.

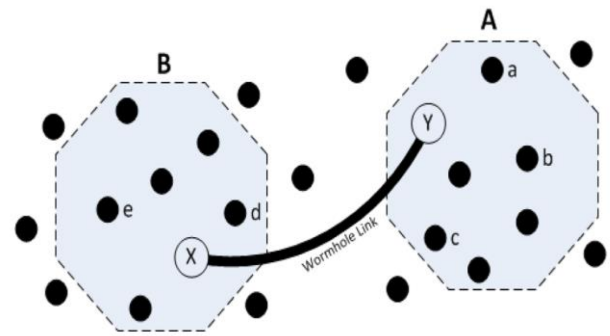


Fig. 5. wormhole attacks In AODV

B. Black hole attack

black hole attack, a malicious node uses its routing protocol in order to with the release of false news, having the shortest path to the destination node or to the packet it wants to avoid the. This black hole node advertises its availability of fresh routes irrespective of checking its routing table. in the attacker node will always have the availability in replying to the route request and thus intercept the data packet and retain it [28]. In protocol based on flooding, the black hole node reply will be received by the requesting node before the reception of reply from actual node; hence a black hole and forged route is created. When this route is establishing, now it’s up to the node whether to drop all the packets or forward it to the unknown address. The Solution how black hole node Proportional in the data routes varies. Fig. 6 shows how black hole Problems, here node “E” want to send data packets to destination node “D” and The initial process of route discovery. So if node “F” is a black hole node then it will claim that it has active route to the specified destination as soon as it receives RREQ packets. It will then send the response to node “E” before any other node. In this way node “E” will think that this is the active route and thus active route discovery is complete. Node “E” will ignore all other replies and will start seeding data packets to node “F”. In this way all the data packet will be lost consumed or lost.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

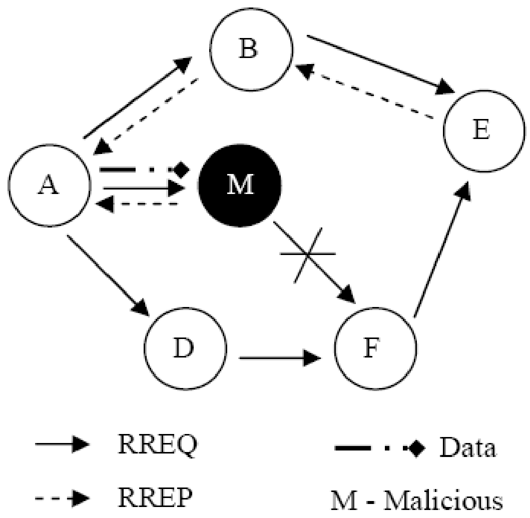


Fig. 6. Black hole attack in AODV

VI. PROPOSED METHOD

In this Paper we represent a new against Black hole and Wormhole Attack in Mobile Ad-Hoc Networks based On Hop Count and RTT way called the protection of navigational protocol in occasional portable networks against black holes, and worm holes which is consisted of, hop count & RTT. The proposed method not only specifies forged ways but also it adopts prevention criteria against reborn of destructive attacks in the process of route detection phase.

A. Second Phase: prevention of black holes, wormholes incidence, between the inception and destination with the usage hop count

In ad-hoc On-demand distance vector (AODV) that is a routing protocol in which each node maintains a routing table, one entry per destination, which records the next hop to the destination and its hop count. AODV also uses a sequence number to ensure freshness of routes. AODV discovers a route through network-wide broadcasting. The RREQ ID field is incremented by one from the last RREQ ID used by the current node. Each node maintains only one RREQID. The Hop Count field is set to zero. When a node receives a RREQ, it first creates or updates a route to the previous hop without a valid sequence number then checks to determine whether it has received a RREQ with the same Originator IP Address and RREQ ID within at least the last If such a RREQ has been received, the node silently discards the newly received RREQ. The rest of this subsection describes actions taken for RREQs that are not discarded. In this proposed method, all existing route to the destination node if the node is examined and the number of hop count have fewer And in the way the company is well-known and have lower energy side of the path, and we do not use this route consider as Malicious path of the cycle cast aside as Next, open the node that has less battery power and the number of hop count have low hop count, but the most responsive to a source node sends this node as malicious node is considered and the cycle of operations being cast aside and The information in this node to all neighboring nodes are distributed. And hop count number to zero, and the path we choose will have more battery power and is submitted in response to any request path number is a good step.

B. The secondly phase: calculating the distance between the inception and destination with RTT

This mechanism which is based on the time and table is used for specifying attacks concerning black holes, silver holes and worm holes, regarding in navigational protocol in the occasional networks, destructive loop always respond to rout reply as quickly as possible and because of this fact that operative loops or the inception doesn't have any valid information about its distance from the destination, they are deceived rapidly by response of these loops, so for preventing from happening this, we must calculate the total round-trip time for all ways available, between the inception and the destination loop until the actual distance between the inception and destination loop is determined and we can make an accurate decision for transmitting information.

1) Calculating round-trip time or RTT:

Round-trip time; middle interval; can be calculated when the inception loop retransmitted "HELLO" message and the moment the response message for "Hello" is received, meaning it has been informed of the neighbor's existence. Each individual loop reserves mono-step round-trip time between itself and its neighbors. Aggressor detection mechanism:
A typical loop which have a data for transmitting, should at first detect necessary route toward the destination and after that start transmitting data using that route. This method is exclusively used in specifying safe route.

2) Aggressor detection mechanism:

A typical loop which have a data for transmitting, should at first detect necessary route toward the destination and after that start transmitting data using that route. This method is exclusively used in specifying safe route.
In order to do this, the data transmitting loop perform the following mathematics:

- For each available or detected route, its RTT is calculated from the inception to destination.
- Then the total amounts RTT for all discovered routes is calculated.
- Calculating the RTT's average for all discovered routes using the previously mentioned amounts.
- Now between all available routes, we choose a route that its RTT has a bigger differentiation, in relation to

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Simulator	NS2.35
Area	500m X500m
Number OF Mobile Node	150
Routing Prptocol	(AODV,DSR)
Transmision Range	250m
Antenna	OmniAntenna
Simulation duration	100,200,300,400,500,600
MAC Layer	802_11
Traffic Type	CBR
Buffer Size	50 Packet
Node placement	Random
Attacker node	64

average RTT amount and utilizing that safe route, we send data toward the destination. According to equation (1)

$$\text{Sending time} = \text{value of receive timestamp} - \text{value of original timestamp} \quad (1)$$

Equation (1) specifies transmitting time course which in fact is a timestamp, representing the differentiation between receiving timestamp with transmitting timestamp.

$$\text{Receiving time} = \text{time the packet returned} - \text{value of transmit timestamp} \quad (2)$$

Equation (2) specifies receiving time course which in fact is a timestamp, representing the differentiation between transmitting timestamp and receiving timestamp.

$$\text{Round-trip time} = \text{Sending time} + \text{Receiving time} \quad (3)$$

In equation (3) we can see round-trip time or RTT. In fact, the total time is the sum of receiving time and sending time.

As a result, when the RTT is calculated for all routs between the inception and destination, the inception loop can easily make an accurate decision regarding its distance from destination and prevents from wrong information, constructed by wormhole loops. Therefore, by applying points which have been mentioned previously, we specify the destructive loops and wipe them out from circuit.

VII. COMPARE THE PROPOSED METHOD (P-METHOD) WITH DSR AND AODV UNDER ATTACK

The reason is that the process of route discovery in DSR and AODV is performed in an on-demand method and nodes tend to discover routes once they require a new route. Since route discovery in DSR and AODV starts and continues by broadcasting RREQ packets, it is much likely that RREQ packets become received by the attacker node. In this case, attacker creates and sends forgery RREP packets to the source node. Such a route advertisement will be much more effective by misusing overheard RREPs and suggesting more fake routes under New Black Hole and worm hole in DSR and AODV algorithm. we used the RTT and Hop count was applied to select the best, optimum route. two parameters including (hop counts and RTT) and source and destination nodes distances were applied in this work to select the optimum route, whereas through the standard DSR and AODV, only RREP and RREQ criterion is applied. The simulation results in section 8 show that Performance of the proposed method in comparison whit DSR and AODV under attack is better.

TABLE.1. SIMULATION PARAMETRS

VIII. ANALYSIS SIMULATION EXPERIMENTS OF TWO ATTACK PATTERNS

This section illustrates how we carry out simulation experiments of the RREQ flooding attacks, the passive black hole attacks and the active black hole attacks, on AODV and DSR protocol using the Network Simulator 2 (ns2.35). Based on the experiment result, we analyze the impact of two attacks on network performance.

A. Experiment Method

Simulation and evaluation of Accurate Black hole attack. and worm hole Compared with AODV under attacks We compared the performance of AODV and DSR routing protocols with black hole attack against the performance of the routing protocols without Black hole attacks. With the help of the Network simulator ns-2 version 2.35 we were able to prove, ns is simulator project, we run two simulations, one AODV and DSR under attacker node and other including the Defense Against Black hole and worm hole Attacks, we have repeated the experiments by changing the Several times. To, 100,200,300,400,500, And 600 to see the simulation parameter are show in table 1 the metrics used to evaluate the performance are given below.

B. JITTER

Jitter is an undesirable effect caused by the inherent tendencies of TCP/IP networks and components. This topic describes the cause and effect of jitter. Jitter is defined as a variation in the delay of received packets. The sending side transmits packets in a continuous stream and spaces them

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

evenly apart. Because of network congestion, improper queuing, or configuration errors, the delay between packets can vary instead of remaining constant,

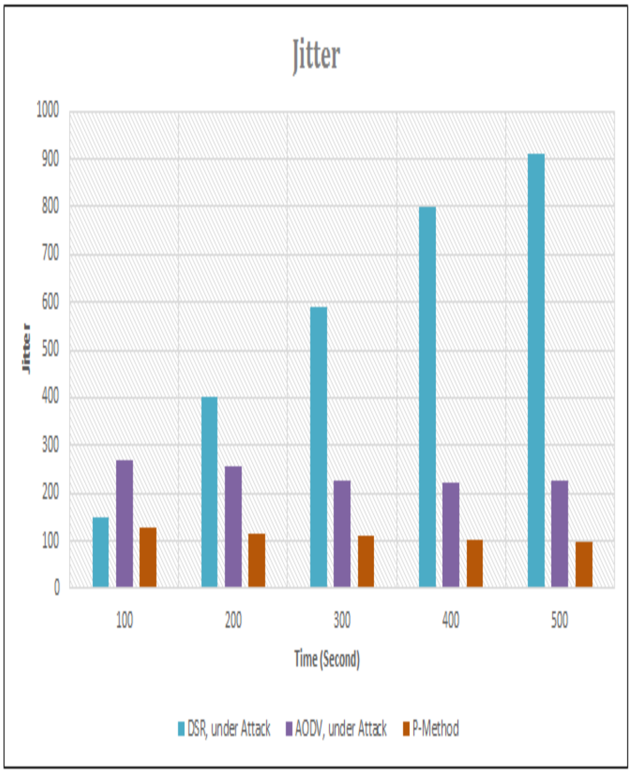


Fig. 7. jitter vs simulation time

Figure 7. shows that the Jitter of Proposed method at the different time Than the DSR under attack and AODV under attack is better, the Jitter at the time of 300 to 500, is better Because proposed method (P-method) is well trained with two phases (Hop Count & RTT).

C. Throughput

a network can be measured by using the different tools that are available on the different operating systems. This page explains the theory, on which the adjustments of these tools for measurements are based, and the issues related to these measurements. The reason for measurement of the throughput in networks is that, the people often intend to know about the maximum operational power of data in a connection link or network access as expressed by the unit of bit per second. The measurement of this quantity is commonly carried out by transmitting a large size file from one system to another and calculating the required time for complete transmitting or copy of the file. Then, with dividing the file size by that time, the throughput will be achieved in unit of megabit per second, kilobit per second or bit per second. The following formula shows how to calculate the throughput

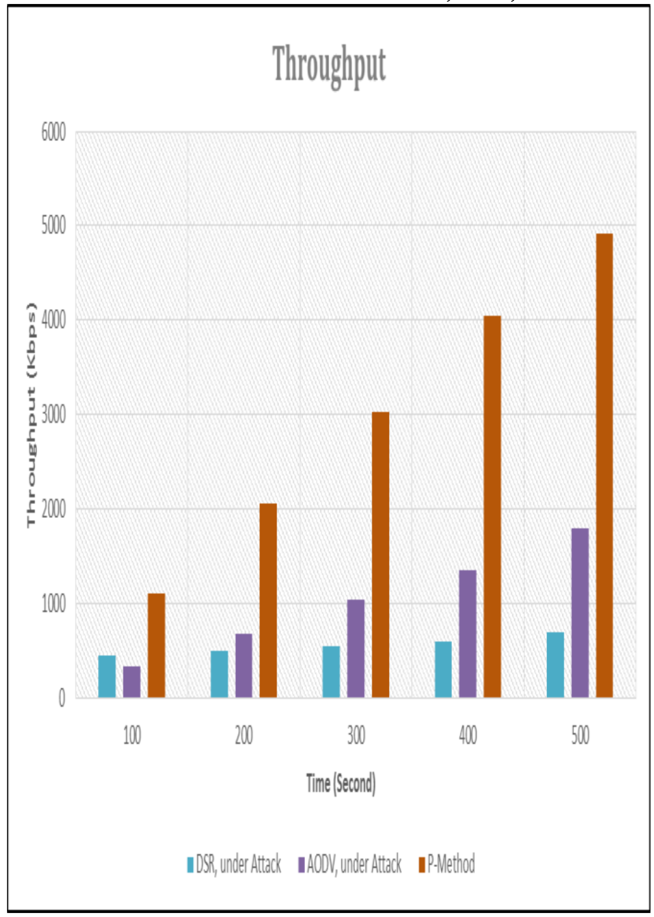


Fig. 8. throughput vs simulation time

Figure 8. Show that for 200 to 500 time, it is obvious that the throughput for Propose method (P-method) is high compared to that of DSR under attack and AODV Under attack. As throughput is the ratio of the total data received from source to the time it takes till the receiver receives the last packet. The overall low throughput of DSR under attack and AODV under attack is due to route reply. the black hole node immediately sends its RREP and the data is sent to the black hole node which cast off all the data. The network throughput is much lower. This result reflects that propose method (P-method)our detection is valid for Defense Against black hole attack and worm hole attack.

D. End to End Delay

In Mobil Ad hoc network, End-to-end delay refers to the time taken for a packet to be transmitted Around the Network from source to destination

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

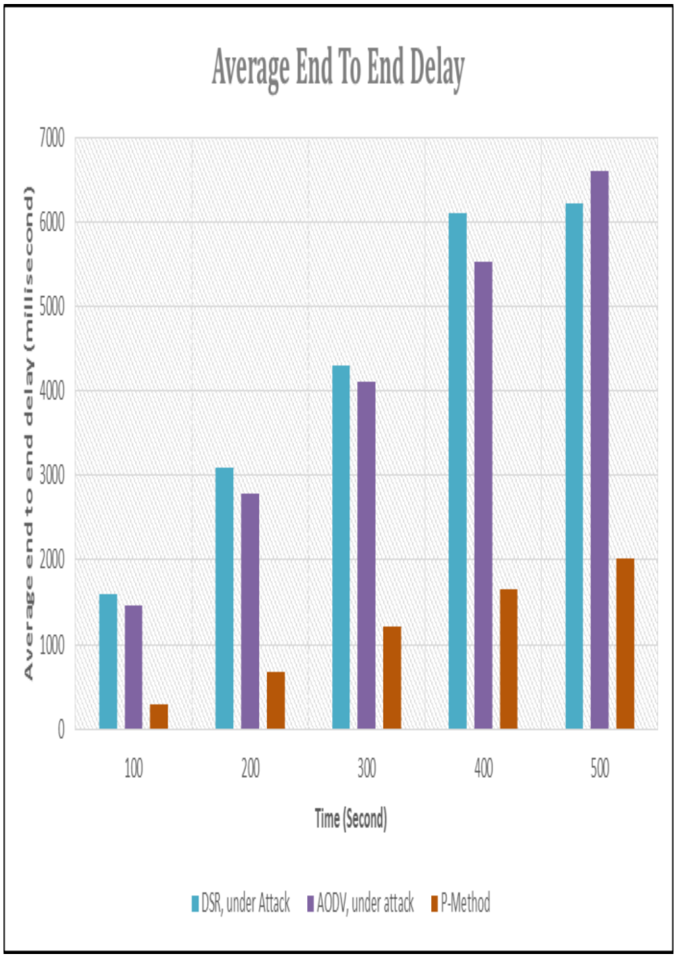


Fig. 9. End to End Delay vs simulation time

Figure 9 show that end-to-end delay is DSR under attack and AODV under attack Considerably higher. compared to Proposed method (P-method). This result reflects that our detection method is valid for Defense Against black hole attack at different times. we used the two phase (RTT and Hop count).

E. Packet Drop ratio (%)

In mobile ad hoc network, a packet drop attack or black hole attack is a type of denial-of-service attack in which a router that is supposed to relay packets instead discards them. This usually occurs from a router becoming compromised from a number of different causes. Because packets are routinely dropped from a loss network.

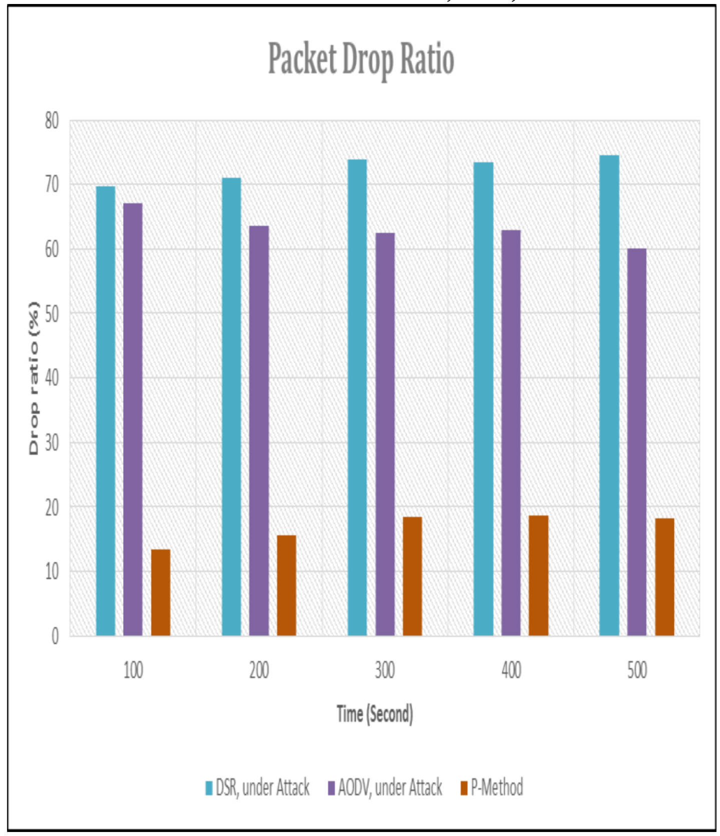


Fig. 10. packet dropped vs simulation time

Figure 10. shown that Black hole has dramatically The drop packet, ratio compared to proposed method, show that in DSR under attack and AODV under attack higher drop packet at different time. This propose method (P-method) result reflects that our detection is valid for Defense Against black hole attack at different times.

IX. Conclusion

In this paper, the DSR and AODV protocol was applied to study selecting the optimum route among the available routes during mobile ad hoc networks routing process. Therefore, the RTT and Hop count was applied to select the best, optimum route. two parameters including (hop counts and RTT) and source and destination nodes distances were applied in this work to select the optimum route, whereas through the standard DSR and AODV, only RREP and RREQ criterion is applied. The simulation results (P-method) Using two phases (Hop count and RTT) Our approach is optimal, Simulation results show that the improved AODV protocol has a distinct advantage in terms of throughput, Jitter, packet dropped, average End to End delay.

Refrence

1. K. Biswas and Md. Liaqat Ali, "Security threats in Mobile Ad-Hoc Network", Master Thesis, Blekinge Institute of Technology" Sweden, 22nd March 2007

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

2. G. A. Pegueno and J. R. Rivera, "Extension to MAC 802.11 for performance Improvement in MANET", Karlstads University, Sweden, December 2006
3. S. Lu, L. Li, K.Y. Lam, L. Jia, "SAODV: A MANET Routing Protocol that can Withstand Black Hole Attack.," International Conference on Computational Intelligence and Security, 2009.
4. M.T.Refaei, V.Srivastava, L.Dasilva, M.Eltoweissy, "A Reputation-Based Mechanism for Isolating Selfish nodes in Ad-Hoc Networks," Second Annual International Conference on Mobile and Ubiquitous Systems, Networking and Services, pp.3-11, July, 2005.
5. S. Lu, L. Li, K.Y. Lam, L. Jia, "SAODV: A MANET Routing Protocol that can Withstand Black Hole Attack.," International Conference on Computational Intelligence and Security, 2009.
6. K. Biswas and Md. Liaqat Ali, "Security threats in Mobile Ad-Hoc Network", Master Thesis, Blekinge Institute of Technology" Sweden, 22nd March 2007
7. S.Marti, T.J.Giuli, K.Lai, M.Baker, "Mitigating Routing Misbehavior in Mobile Ad-Hoc Networks".
8. M.T.Refaei, V.Srivastava, L.Dasilva, M.Eltoweissy, "A Reputation-Based Mechanism for Isolating Selfish nodes in Ad-Hoc Networks," Second Annual International Conference on Mobile and Ubiquitous Systems, Networking and Services, pp.3-11, July, 2005.
9. K. Biswas and Md. Liaqat Ali, "Security threats in Mobile Ad-Hoc Network", Master Thesis, Blekinge Institute of Technology" Sweden, 22nd March 2007
10. S.Marti, T.J.Giuli, K.Lai, M.Baker, "Mitigating Routing Misbehavior in Mobile Ad-Hoc Networks".
11. M.T.Refaei, V.Srivastava, L.Dasilva, M.Eltoweissy, "A Reputation-Based Mechanism for Isolating Selfish nodes in Ad-Hoc Networks," Second Annual International Conference on Mobile and Ubiquitous Systems, Networking and Services, pp.3-11, July, 2005.
12. N.Bhalaji, Dr.A.Shanmugam, "Association between nodes to combat blackhole attack in DSR based MANEr", Proc. Int. IFIP Conf. on Wireless and Optical Communications Networks, WOCN '09, India, pp. 1-5,2009.
13. V. Palanisamy, P. Annadurai, S.Vijayalakshmi ,"Impact of Black hole Attack on Multicast in Ad hoc Network (IBAMA)", Proc. In!. IEEE Conf. on Computational Intelligence and Computing Research (ICCIC), India, pp. 1-4,2010.
14. Sh. Sharma, R. Gupta,"Simulation study of Blackhole attack in the mobile ad hoc networks", Journal of Engineering Science and Technology Vol. 4, No. 2 (2009) P243 - 250 © School of Engineering, Taylor's University College.
15. K. Biswas and Md. Liaqat Ali, "Security threats in Mobile Ad-Hoc Network", Master Thesis, Blekinge Institute of Technology" Sweden, 22nd March 2007
16. G. A. Pegueno and J. R. Rivera, "Extension to MAC 802.11 for performance Improvement in MANET", Karlstads University, Sweden, December 2006
17. Latha Tamilselvan, V. Sankaranarayanan: "Prevention of Black Hole Attack in MANEr", The 2nd international conference on wireless, Broadband and Ultra Wideband Communications (January 2007)
18. Mohanmmad AI-Shurrnan et al: "Black Hole Attack in Mobile Ad Hoc Network", ACMSE' 04, (April 2004)
19. A Trusted AODV Routing Protocol for Mobile Ad Hoc Networks. PhD thesis, Department of Computer Science and Engineering, The Chinese University of Hong Kong, 2003
20. C.E. Perkins, S.R. Das, and E. Royer, "Ad-Hoc on Demand Distance Vector (AODV)", March 2000, http://www.ietf.org/internet-drafts/draft-ietf-manet-aodv- 05.txt
21. Charles E. Perkins, Elizabeth M. Belding-Royer, Samir R. Das, Mobile Ad Hoc Networking Working Group, Internet Draft, 17 February 2003.
22. D.Johnson, Y. Hu, and D. Maltz, The Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4, RFC 4728, 2007.
23. C.E. Perkins, S.R. Das, and E. Royer, "Ad-Hoc on Demand Distance Vector (AODV)", March 2000, http://www.ietf.org/internet-drafts/draft-ietf-manet-aodv- 05.txt
24. Charles E. Perkins, Elizabeth M. Belding-Royer, Samir R. Das, Mobile Ad Hoc Networking Working Group, Internet Draft, 17 February 2003.
25. Azer, M.A., El-Kassas S.M., Hassan, A.W.F., El-Soudani M.S., "Intrusion Detection for Wormhole Attacks in Ad hoc Networks a Survey and a proposed Decentralized Scheme Marianne " IEEE Third International conference on Availability, Reliability and Security, 2008.
26. Reshmi Maulik, Nabendu Chaki "A comprehensive review on wormhole attacks in MANET" International Conference on Computer Information Systems and Industrail Management Applications(CISIM) 2010.
27. K. Biswas and Md. Liaqat Ali, "Security threats in Mobile Ad-Hoc Network", Master Thesis, Blekinge Institute of Technology" Sweden, 22nd March 2007
28. G. A. Pegueno and J. R. Rivera, "Extension to MAC 802.11 for performance Improvement in MANET", Karlstads University, Sweden, December 2006
29. Reza Fotohi, Shahram Jamali, Fateme Sarkohaki, Shahram Behzad,"An Improvement over AODV Routing Protocol by Limiting Visited Hop Count", IJITCS, vol.5, no.9, pp.87-93, 2013. DOI: 10.5815/ijitcs.2013.09.09

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 6, June 2016

Check The Use of Raise in wireless sensor networks based on heuristic algorithms along with soft computing approach

[1]*Abolfazl akbari* — Department of Computer Engineering, Ayatollah Amoli Branch, Islamic azad university, Amol,Iran

[2] *poorya khodabandeh* — Department of Computer Engineering, Marlik Higher Education Institute, nowshahr, Iran

[3]*ali khosrozadeh* — Department of Computer Engineering, Ayatollah Amoli Branch, Islamic azad university, Amol,Iran

Abstract

The use of Wireless Sensor Networks (WSNs) has grown dramatically in recent decades, and the use of these networks in the areas of military, health, environment, business, etc. increases every day. A wireless sensor network consists of many tiny sensor nodes with wireless communications and work independently. In applications of such sensor nodes, hundreds or even thousands of low-cost sensor nodes are dispersed over the monitoring area, in which each sensor node periodically reports its sensed data to the base station (sink). Due to limitations in the communication range, sensor nodes transmit their sensed data through multiple hops. Each sensor node acts as a routing element for other nodes for transmitting data.

One of the most important challenges in designing such networks is the management of energy consumption of nodes; because replacing or charging the batteries of these nodes are usually impossible.

One of the main characteristics of these networks is that the network lifetime is highly related to the route selection. Unbalanced energy consumption is an inherent problem in WSNs characterized by the multi-hop routing and many-to-one traffic pattern. This uneven energy dissipation in many routing algorithms can cause network partition because some nodes that are part of the efficient path are drained from their battery energy quicker. To efficiently route data through transmission path from node to node and to prolong the overall lifetime of the network, In this thesis we proposed three new routing algorithms using a combination of both Fuzzy approach and A-star algorithm seeks to investigate the problems of balancing energy consumption and maximization of network lifetime for WSNs :A-Star with 3 parameters fuzzy system (A*3F), A-Star with 3 fuzzy system with 2 parameters using majority vote (A*3FMV) and A-Star with 3 fuzzy system with 2 parameters using simple additive weighting (A*3FSAW). The new methods is capable of selecting optimal routing path from the source node to the sink by favoring the highest remaining energy, minimum number of hops, lowest traffic load and energy consumption rate.

We evaluate and compare the efficiency of the proposed algorithms with each other methods under the same criteria in four different topographical areas. Simulation results show that A*3PFSAW and A*3PFMV balances the energy consumption well among all sensor nodes and achieves an obvious improvement on the network lifetime that randomly scattered nodes and flat routing..

Keywords: Wireless Sensor Networks, A-Star algorithm, Fuzzy logic, Network lifetime, Multi-hop routing.

1. Introduction

A wireless sensor network is a collection of nodes that form a network working together. Each node has a processing capability, memory, a transmitter / receiver RF, a unit of power (battery or solar cell) and can have different types of sensors are operating. After the nodes in a distributed environment, wirelessly communicate with each other and organize themselves into a contingency operation as a whole.

Since sensor networks can contain various types of sensors such as vibration sensor, magnetic, thermal, acoustic, visual and radar, so can monitor the various environmental conditions such as temperature, humidity, movement of vehicles, the lightning, the pressure, noise levels, the presence or absence of certain kinds of objects, mechanical pressure levels on the objects, properties of objects, such as current speed, direction and size.[1]

Sensor nodes can be continuously use for discovery event, a sense of place and local control. Features of micro-sensing and wireless communication between the nodes, promising many applications in the new fields of applications such as fields of military, health, home and business and categorized into the areas of space exploration, chemical treatments and relief for natural disaster.

Usually, sensor nodes are randomly distributed in the environment. The main components of communication are:

• Sensor nodes. Each of these nodes, the ability to collect and send data to the sink, wirelessly. Communicate with the sink nodes can be single-stage or multi-stage.

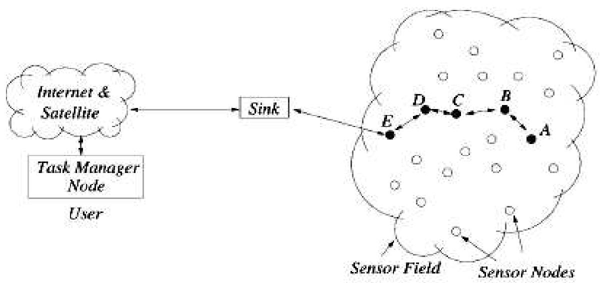


Fig.1. communication architecture for wireless sensor networks

In the past few years, intensive research on the potential of collaboration among sensors to collect and process sensed data and the coordination and management of activities are performed. However, sensor nodes have limited energy supply and bandwidth. Thus, innovative ways to eliminate inefficiencies in energy constraints that reduces the lifetime of the network is required.

Despite the innumerable applications of wireless sensor networks, these networks have several limitations, for example, limited energy supply, limited computing power, and limited bandwidth of wireless links.

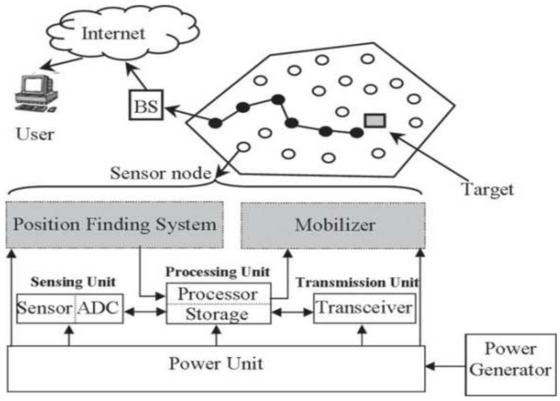


Fig.2. Components of a sensor node

One of the main goals of the design of wireless sensor networks Performing data communication while trying to prolong network lifetime and to prevent damage connection by applying energy management techniques. The design of routing protocols in wireless sensor networks is affected by many challenging factors. Before communication effectively in wireless sensor networks must overcome these factors.Some routing challenges and design issues that affect the routing of wireless sensor networks are deploying nodes, the energy consumption without loss of accuracy of the reported data, the heterogeneity of nodes and connections, fault tolerance, scalability, network dynamics,

• The base station (sink) that communicates with the user via the Internet or satellite.

• Something that user wants to receive information about it.

• The user that data collected to measure / monitor the behavior of the phenomenon.

communication medium, density or density, coverage area, data integration, quality of service and ...[4].

Due to limitations in the communication range, sensor nodes transmit their sensed data through multiple hops. Each sensor node acts as a routing element for other nodes for transmitting data. Energy is therefore a crucial parameter in power-constrained data-gathering sensor networks. Energy consumption should be well managed to maximize the network lifetime [5]. Unbalanced energy consumption is an inherent problem in WSNs characterized by the multi-hop routing and many-to-one traffic pattern. The uneven energy dissipation can significantly reduce network lifetime. Generally in routing algorithm, the best path is chosen for transmission of data from source to the destination. Over a period of time, if the same path is chosen for all communications in order to achieve battery performance in terms of quick transmission time, then those nodes on this path will get drained fast [3], [5], [7].The problem with many algorithms is that they minimize the total energy consumption in the network at the expense of non-uniform energy drainage in the networks. Such approaches cause network partition because some nodes that are part of the efficient path are drained from their battery energy quicker.

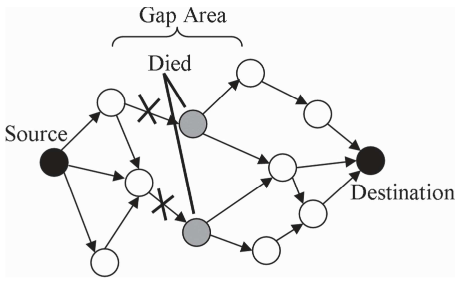


Fig.3.Network partition due to the death of certain nodes

Thefuzzyinferencesystem (FIS) can optimizes therouting path (depending on themetrics: distance, remaining battery powerand energy consumption rate) in a distributed fashion. When a data is needed to be sent the protocol selects the optimal path through the FIS. Designers and developers of protocolsand applications for WSN have emphasized on heuristic search technique, called A-Star algorithm, for searching best path for routing in WSN. They suggest that the criteria to search best path is not only to get path with minimum energy consumption but also to see that no deselected in the path contain enough of resi dual energy.

Therefore, in this paper, the proposed method for balancing energy consumption and maximization of network lifetime for WSNs. We propose a new approach by combining Mixed-Fuzzy approach and A-star algorithm to select the optimal routing path from the source to the destination by favoring the highest remaining battery power,

minimum number of hops ، minimum traffic loads and minimum energy consumption rate.

2. RELATED WORKS

Many challenges are in the design of wireless sensor networks such as energy efficiency, network scalability, network operating environment, the fault-tolerance, data delivery models, data integration, quality of service, delay, distribution of nodes, mobility or lack of mobility of nodes, the nodes are identical or not, network congestion, etc., which is one of the most prominent and important of these challenges, the problem of limited energy and how efficient it is to have a significant impact on how routing and a lot of research in this field is such that it can be cited, such as the following: for example the work in [8] proposed to minimize the hop stretch of a routing path (defined the shortest path) in order to reduce the energy cost of end-to-end transmission. The approaches in [9], [10] took a different view for prolonging the network-lifetime. They attempt to sustain the availability of the sensors that have less energy by distributing the traffic load to the ones with much residual energy. All of the above-mentioned works focus on improving energy-efficiency using fixed routing paths; nonetheless, due to the lack of path diversity, those nodes traversed by fixed routing paths may drain out their energy quickly.

The work in [11] exploited two natural advantages of opportunistic routing, i.e. path diversity and the improvement of transmission reliability, to develop a distributed routing scheme for prolonging the network lifetime of a WSN. The goal of this work is to assist each sensor in determining a suitable set of forwarders as well as their priorities, thus, enabling effort to extend the network-lifetime. Madan et al. in [12] solved the lifetime maximization problem with a distributed algorithm using the dual decomposition and the sub gradient method. Chang and Tassiulas in [13] proposed a shortest cost path routing algorithm for maximizing network lifetime based on link costs that reflect both the communication energy consumption rates and the residual energy levels. The authors of [14] presented a uniform balancing energy routing protocol to choose the nodes whose residual energies were greater than a certain threshold as routers for other nodes in every transmission round, and distributed the energy load among any sensors to maximize the whole network lifetime.

Lu et al. in [15] proposed an Energy-Efficient Multi-path Routing Protocol (EEMRP). It has the capability of searching multiple node-disjoint paths and utilizes a load balancing method to assign the traffic over each selected path. Both the residual energy level of nodes and the number of hops are considered to be incorporated into the link cost function. It uses a fairness index to evaluate the level of load balancing over different multi-paths. Furthermore, since EEMRP only takes care of data transfer delay, the reliability of successful paths sometimes is limited. The authors in [16] presented a new routing protocol based on a high weight genetic algorithm. In this method, the sensor nodes are aware of the data traffic rate to monitor the network congestion.

FML-MP (a fuzzy multi-path maximum lifespan routing scheme), an online multi-path routing scheme that strives to achieve a good distribution of the traffic load is developed in [17]. It uses an edge-weight function in the path search process.

In [18] the authors presented Optimal Forwarding by Fuzzy Inference Systems (OFFIS) for flat sensor networks. The OFFIS protocol selected the best node from candidate nodes in the forwarding paths by favoring the minimum number of hops, shortest path and maximum remaining battery power, etc. The authors in [19] presented a novel algorithm for routing analysis in WSNs utilizing a fuzzy logic at each node to determine its capability to transfer data based on its relative energy levels, distance and traffic load to maximize the lifetime of the sensor networks.

Rana et al. in [20] used A-star algorithm to search optimal route from the source to destination in such a way that, there is a pre-defined minimum energy level for sensor nodes so that sensor node doesn't participate in routing if its residual energy level is below that level.

Deepak S. Gaikwad and SampadaPimpale in [29] and have presented a combination protocol (A-Star with fuzzy) like such we have proposed, major weakness of this protocol considering only two input parameters residual energy level and the traffic load and they considered the time of death the first live nodes in the network without checking history of energy consumption rate at each node as base of improvement , which was summarized in comparison with the proposed protocols can be said that the time of death of the first node , the number of nodes remaining alive at the end of the scenarios and remaining energy in different algorithms are influenced by factors such as geographical location in the network, moving the BS, and the size of the network (length and width) and the network will behave differently, however, in a square network field, the proposed Gaikwad and Pimpale method ,the time of death of the nodes be longer but in our proposed method by proper using of energy consumption rate (ECR) as third parameter for selecting optimal path the network lifetime is prolonged. Even using (ECR) in one of the proposed methods, the number of nodes alive at the end of the scenario which leads to higher levels of residual remaining energy.

In most applications of WSNs, sensor nodes are densely deployed in large areas. Once deployed, nodes can never be recharged or replaced. After depleting their energy, nodes turn to die and stop working. Since networks cannot accomplish assigned missions after nodes die [4], [6]. The maximization of lifetime can be formulated asan optimization problem. The variable sof this optimization problem are routing parameters atnodes. When having sensedor asked to relay a data packet, each node needs to transmit this packet to a sink. However, it cannot send the packet directly to sinks except that it is a sink's neighbor. So normally a node needs to choose a neighboring sensor as its next hop. When nodes are chosen as the next hops they will influence the energy consumption of the network as well as the lifetime.

Energy Balanced Distributing in Routing is one of the solutions for maximize network lifetime and optimized

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

management in energy consumption. WSN networks often suffer from the problem of using uneven energy, the unfavorable energy dissipation causes network lifetime of WSN can be severely reduced.

From the aforementioned literatures, we note that a number of different metrics have been used to prolong the lifetime of the sensor networks such as : *Remaining Energy(RE)* [3],[15], [21], *Minimum Hop (MH)* [15], [18], [19], [21]*and Traffic Load (TL)*[3], [16], [19], [21].

To extend the network lifetime, this paper proposes a new routing method using a combination of Mix-Fuzzy approach and A-star algorithm. The proposed routing method is used to select the optimal routing path from source to destination by considering Remaining Energy, Minimum Hop, Traffic Load and Energy Consumption rate and balancing between them to lengthen the lifetime of the sensor network as much as possible.

3. Fuzzy Approach

Fuzzy logic was first introduced in the mid-1960s by Lotfi-Zadeh in [22]. Since then, its applications have rapidly expanded in adaptive control systems and system identification. It has the advantages of easy implementation, robustness, and ability to approximate to any nonlinear mapping.

Fuzzy logic analyzes information using fuzzy sets, each of which is represented by a linguistic term such as “small,” “medium,” or “large.” Fuzzy sets allow an object to be a partial member of a set. In Fig. 4, if X suggests a collection of objects denoted by x , usually X is referred to as the “universe of discourse,” and then a fuzzy set A in X is defined by a set of ordered pairs:

$$A = \{(x ,\mu_A(x))/x \in X \}. \qquad (1)$$

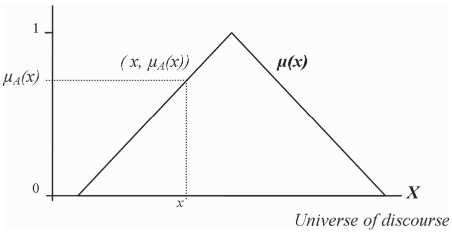


Fig.4.Membership function from the pair (*x, μA(x)*).

Where the function $\mu_A(x)$ is called membership function of the object x in A. This membership function represents a “degree of belongingness” for each object to a fuzzy set, and provides a mapping of objects to a continuous membership value in the interval [0...1]. When a membership value is close to the value 1 ($\mu_A(x) \longrightarrow 1$), it means that input x belongs to the set A with a high degree, while small membership values ($\mu_A(x) \longrightarrow 0$), indicate that set A does not suit input x very well [23].

Infuzzy systems,the dynamic behavior of a system is characterized by a set of linguistic fuzzy rules based on the knowledge of a human expert. Fuzzy rules are of the general form: If antecedent(s) then consequent(s), where antecedent sandconsequents are propositions containing linguistic variables. Antecedents of a fuzzy rule form acombination of fuzzy sets through the use of logic operations.Fig 5 shows the typical structure of a fuzzy system. It consists of four components namely; fuzzification, rule base, inference engine and defuzzification. The processes of making crisp inputs are mapped to their fuzzy representation in the process called fuzzification. This involves application of membership functions such as triangular, trapezoidal, Gaussian etc. The inference engine process maps fuzzified inputs to the rule base to produce a fuzzy output. A consequent of the rule and its membership to the output sets are determined here. The defuzzification process converts the output of a fuzzy rule into crisp outputs by one of defuzzification strategies. Thus, fuzzy sets and fuzzy rules together form the knowledge base of a rule-based inference system.

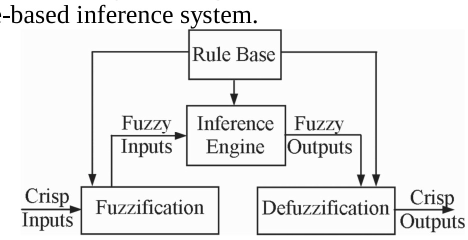


Fig.5.Typical structure of the fuzzy approach

.

Considering a fuzzy system with p inputs and one output with M rules, then the L_{th} rule has the form [22], [23], [24] :

$$IF\ x_1\ is\ f_1^L\ and \ldots xp\ is\ f_L^P \rightarrow Then\ y\ is\ G^L$$

4. A-Star Algorithm

A-star search algorithm is a widely used graphic searching algorithm. It is also a highly efficient heuristic algorithm used in finding a variable or low cost path. It is considered as one of the best intelligent search algorithms that combines the merits of both depth-first search algorithm and breadth-first algorithm.

A-star path searching algorithm uses the evaluation function (usually denoted f (n)) to guide and determine the order in which the search visits nodes in the tree. The evaluation function is given as: $f(n)=g(n)+h(n)$

where g(n) is the actual cost from the initial node (start node) to node n (i.e. the cost finding of optimal path), h(n) is the estimated cost of the optimal path from node n to the target node (destination node), which depends on the heuristic information of the problem area [25].

Generally, A-star algorithm maintains two lists, an OPEN list and a CLOSE list. The OPEN list is a priority queue and keeps track of the nodes in it to find out the next node with least evaluation function to pick. The CLOSE list keeps track of nodes that have already been examined. Initially, the OPEN list contains the starting node. When it iterates once, it takes the top of the priority list, and then checks whether it is the goal node (destination node). If so, the algorithm is done. Otherwise, it calculates the evaluation function of all adjacent nodes and adds them to the OPEN list. After the A-star algorithm is completed, it will find a solution if a solution exists. If it doesn’t find a solution, then it can guarantee that no such solution exists. A-star algorithm will find a path with the lowest possible cost. This will

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

depend heavily upon the quality of the cost function and estimates provided [26].

A-star algorithm (Pseudo-code A*) may be expressed as following [25], [27]:

```
Create the open list of nodes, initially containing only our starting node
    Create the closed list of nodes, initially empty
    While (we have not reached our goal) {
        Consider the best node in the open list (the node with the lowest f value)
        If (this node is the goal) {
            Then we're done
        }
        Else {
            Move the current node to the closed list and consider all of its neighbors
            For (each neighbor) {
  if (this neighbor is in the closed list and our current g value is lower) {
                    Update the neighbor with the new, lower, g value
                    Change the neighbor's parent to our current node
                }
                Else if (this neighbor is in the open list and our current g value is lower) {
                    Update the neighbor with the new, lower, g value
                    Change the neighbor's parent to our current node
                }
                Else this neighbor is not in either the open or closed list {
                    Add the neighbor to the open list and set its g value
                }
            }
        }
    }.
```

5. Simple Additive Weighting Methods of Multi Criteria Decision Making

Various multi-criteria decision making (MCDM) methods have been proposed to solve diverse applications of decision problems. One of the MCDM methods is additive weighting-based method.Simple Additive Weighting (SAW) which is also known as weighted linear combination or scoring methods is a simple and most often used multi attribute decision technique. The method is based on the weighted average. An evaluation score is calculated for each alternative by multiplying the scaled value given to the alternative of that attribute with the weights of relative importance directly assigned by decision maker followed by summing of the products for all criteria. The advantage of this method is that it is a proportional linear transformation of the raw data which means that the relative order of magnitude of the standardized scores remains equal. Process of SAW consist of these steps:

1. Create a Decision Matrix according below table and Quantification of Decision Matrix:

Create a decision matrix table of the output fuzzy systems 1, 2 and 3 according to the following formula is obtained by replacing the values of the output value aij in the matrix multi-criteria fill in. Options include a list of all neighbors of a sensor node is a matrix of which one is selected in the list by the SAW algorithm and the rest of stay in open list. In our proposed methods uses two methods SAW and Majority Vote to decide three expert systems as follows. In the Majority Vote approach between the votes obtained from expert systems, which one that have the highest value is used to jump as destination node and other nodes are in the open list of A-Star algorithm. In SAW, by calculating the indexes weights, contributed to the decision. SAW due to mathematical models have higher accuracy compared to the Majority Vote.

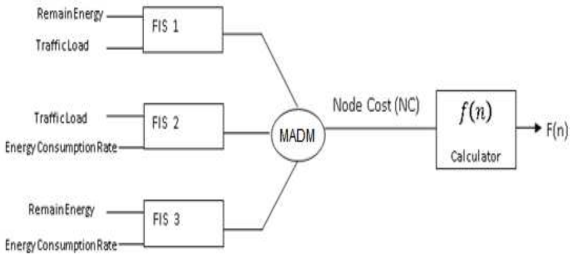


Fig. 6. A*3PFSAW and A*3PFMV methods of three fuzzy systems with two-parameter mixed by MADM

$$f(n) = \frac{1}{\text{Min } Hop\ Node(n)\ to\ BS} + NC(n)$$

n: list of neighboring nodes of a sensor node

Table 1: Multi-Attribute Decision Making Matrix of Scenarios

List of neighbors of a sensor node	Output of fuzzy systems			
	Cj / Ai	FIS 1	FIS 2	FIS 3
	A	$f_A(n)_1$	$f_A(n)_2$	$f_A(n)_3$
	B	$f_B(n)_1$	$f_B(n)_2$	$f_B(n)_3$
	C	$F_C(n)_1$	$F_C(n)_2$	$f_C(n)_3$
	⋮	⋮	⋮	⋮

2. Making bi-linear scaling of the values of the Decision Matrix:

For positive indicators: $n_{ij} =$ For negative indicators: $n_{ij} = 1 -$

For both positive and negative indicators: $n = \frac{\frac{1}{a_{ij}}}{Max(\frac{1}{a_{ij}})} =$

3. Multiplying the matrix of weights and measures Scale:

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$p_{ij} = \frac{a_{ij}}{\sum_{i=1}^{m} a_{ij}} \quad ; \quad \forall_j$$

$$E_j = -k \sum_{i=1}^{m} [\, p_{ij} \ln p_{ij} \,] \quad ; \quad \forall_j \quad k = \frac{1}{\ln(m)}$$

$$d_j = 1 - E_j \quad ; \quad \forall_j$$

$$w_j = \frac{d_j}{\sum_{j=1}^{n} d_j} \quad ; \quad \forall_j$$

4. Choose the best option (A*) using the following criteria:

$$A^* = \{A_i | Max \sum_{j=1}^{n} n_{ij} w_j\}$$

6. PROPOSE D ROUT ING METHOD

In this paper, the topology of a WSN is modeled as a directed graph G(N , A), where N is the set of nodes, and A is the set of direct links between the nodes. A sink node is responsible for collecting data from all other nodes within its transmission range [5], [9], [10], [26]. The routing schedule is computed by the base station. It calculates optimal routing schedule and broadcasts it. Every node follows this schedule. The process of finding the optimal path, and broadcasting it in the network and sending data from all nodes to the base station by following this routing schedule is repeated in every round. Computation of routing schedule is done dynamically with the consideration of current level of some criteria of each node. For this, normally it may require the nodes to report their criteria periodically to the base station. The base station can then determine the routing schedule based on this updated information.

For the proposed model, whenever any sensor node runs out of energy, communication links between various sensor nodes and the base station will break. This is considered as the end of the network lifetime. Since the lifetime of each sensor node depends on energy consumption, it is important to preserve residual energy of these nodes in such a way that overall network lifetime is extended.

To achieve this goal we propose innovative methods and some of these methods to evaluate the efficiency will be compared. first method is only A-Star algorithm alone which is used as the base routing task , in this way does not consider value of parameters like remaining energy , traffic load and energy consumption tare to select neighbor to jump. In this routing method, the base station prepares the routing schedule and broadcast it to each node. A-star algorithm which is used to find the optimal route from the node to the base station is applied to each node. A-star algorithm creates a tree structure in order to search optimal routing path from a given node to the base station. The tree no deisexplored based on it s*evaluation function f(n)*.The function we use disgivenas: f (n) = g(n) + h(n).

The second method combines fuzzy methods and the A-star. The tree node is explored based on its evaluation function f (n). The function we used is given as: *f(n)=NC(n)+(1⁄MH(n))*.Where NC (n) is the node cost of node n, which takes value [0…1], and can be calculated by the fuzzy approach. The fuzzy approach is considered for the remaining energy and the traffic load of node n to calculate the optimal cost for node n. MH (n) is the short distance from node n to the base station. As a result, the node n that has largest f (n) value will be chosen as the optimal node.

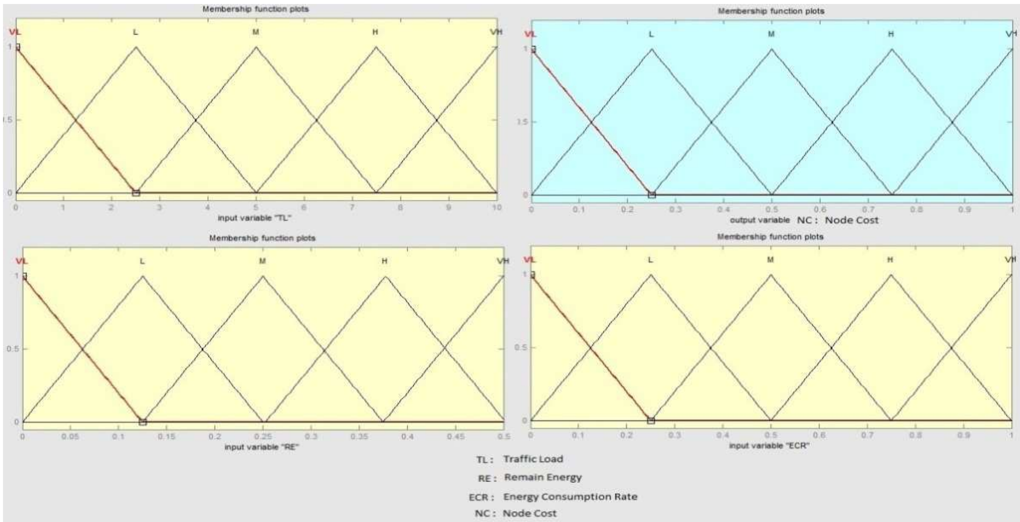


Fig.7.Membership graph for the inputs (remaining energy, traffic load and energy consumption rate) and the output (nodecost).

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

The goal of the fuzzy part of the proposed protocol is to determine the optimal value of the node cost NC(n) of node n that depends on the remaining energy RE(n) and the traffic load TL(n) of node n. Fig. 8 shows the fuzzy approach with two input variables RE(n) and TL(n), and an output NC(n), with universal of discourse [0…5], [0…10], and [0…1], respectively. This method uses five membership functions for each input and an output variable, as shown in Fig. 7.For the fuzzy approach, the fuzzified values are processed by the inference engine, which consists of a rule base and various methods to inference the rules. The rule base is simply a series of IF-THEN rules that relate the input fuzzy variables and the output variable using linguistic variables each of which is described by fuzzy set and fuzzy implication operator AND. Table A in appendix shows the IF-THEN rules used in the proposed method .All these rules are processed in a parallel manner by a fuzzy inference engine. At the end, the defuzzification finds a single crisp output value from the solution fuzzy space. This value represents the node cost. Practice defuzzification is done using center-of-gravity method [24] given by:

$$node\ cost = \frac{\sum_{i=1}^{n} u_i * c_i}{\sum_{i=1}^{n} u_i}$$

Where Ui is the output of rule base i , and ci is the center of the output membership function.

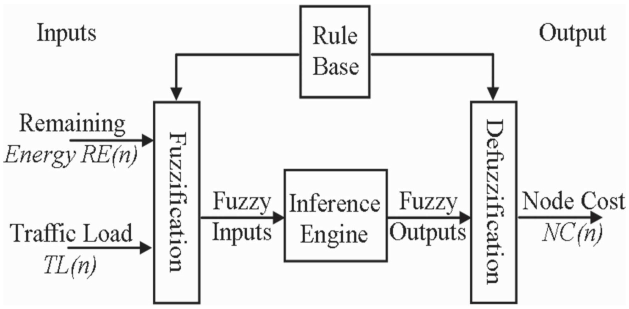


Fig.8.Fuzzy structure with two inputs (remaining energy and traffic load) and one output (nodecost).

A third way to assess the impact of using more parameters in an expert system to decide the choice of the next node in the optimized routing operation uses the third parameter as input of the fuzzy system called the energy consumption rate of the neighbor nodes (ECR). This parameter indicates the use of a node in routing process .The more use of nodes in routing process, the less will be used in the selection of preferred nodes as the relay node. Fuzzy rules are attached to the paper. The flowchart of this new method is similar to flowchart of the second method except that the third arguments ECR have also been used in fuzzy systems.

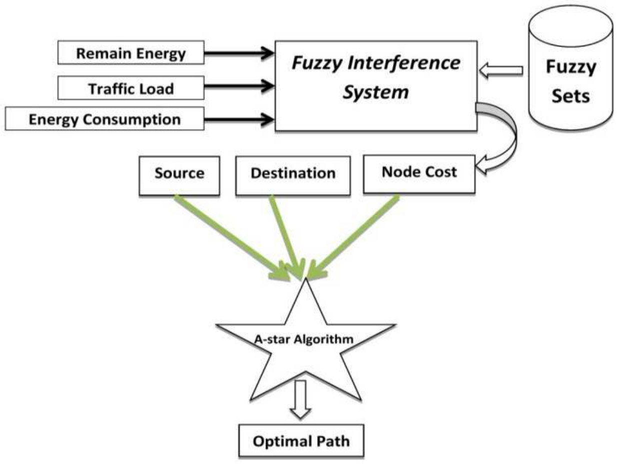


Fig. 9. Two-parameter and three-parameter hybrid algorithm with the A-Star

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

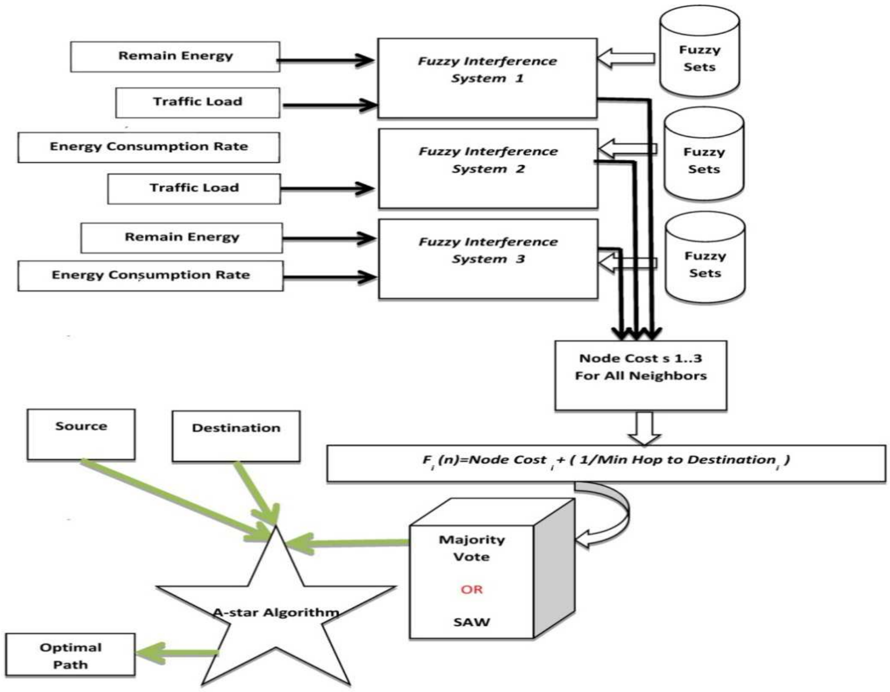


Fig.10. Structure of the other proposed algorithm for routing fuzzy with mixing method of expert systems (SAW or Majority Vote)

The fourth and fifth proposed methods use several fuzzy expert systems (below chart) by applying Majority vote and SAW methods. In this methods aims to enhance the accuracy of the decision making in choosing the optimal path. In this methods after calculating the minimum number of jumps to the destination (MH) by A-star algorithm, then node cost value calculating by three fuzzy expert systems according to the following chart with the corresponding parameters. Then the values use for calculate F(n) for all neighboring nodes which candidate to jump in order to produce multi-criteria matrix after that SAW algorithm use this matrix to selection optimal neighbor. The Majority Vote approach, instead of a creating multi-criteria matrix to select the best neighbor node for next hop use majority vote Fuzzy expert systems, neighbor node with highest vote will be selected.

7. PERFORMANCE EVALUATION

To demonstrate the effectiveness of the proposed methods in terms of balancing energy consumption and maximizing network lifetime, simulation results of the proposed methods are compared with those of A-star search algorithm and with those of Fuzzy mix with A-star (A*2PF) approach and with new methods : Fuzzy 3 parameters mix with A-star (A*3PF) and with three-2 parameters Fuzzy systems mix with A-star by Majority Vote (A*3PFMV) and with three-2 parameters Fuzzy systems mix with A-star by SAW (A*3PFSAW), for four different topographical areas according to Table 2.

The simulations are carried out in MATLAB. 100 sensor nodes are randomly deployed in a topographical area A ,B ,C ,D of dimension 100 m × 100 m and dimension 200 m × 50 m. All topographical areas have the sensed transmission limit of 30 m. The performance of the proposed method is tested in these four topographical areas. There is only one data sink which located at (90 m, 90 m) for area A and at (180 m, 45 m) for area B. All sensor nodes have the same initial energy 0.5 J . The proposed method uses the W.R.Heinzelman radio model that is largely used in the area of routing protocol evaluation in WSNs [28].

$$E_{Tx}(k,d) = E_{Tx-elec}(k) + E_{Tx-amp}(k,d) = E_{elec} * k + \varepsilon_{amp} * k * d^2$$
$$E_{Rx}(k) = E_{Rx-elec}(k) = E_{elec} * (k)$$

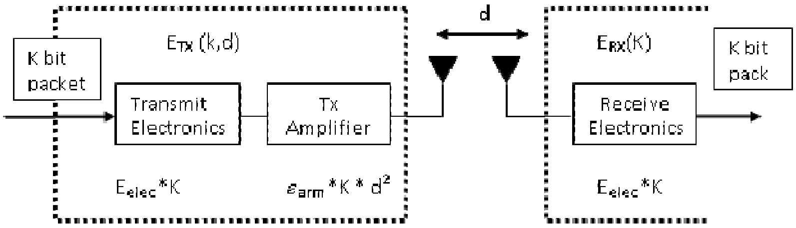


Fig .11. Diagram of the transmitter and receiver wireless sensor

According to this model, transmission and receiving costs are characterized by the expressions $E_{Tx}(k,d)$ and $E_{Rx}(k)$, respectively, where k is the number of bit per packet, d is the distance from the sender node to the receiver node, E_{elec} and E_{amp} are per bit energy dissipation in transmitting or receiving circuitry and energy required per bit per meter square for the amplifier to achieve acceptable signal to noise ratio (SNR) respectively. Simulations are done using the values 50 nJ/bit and 100 pJ/bit/m2 for E_{elec} and E_{amp} , respectively. The traffic load, in each node is assumed to be generated randomly between [0...10]. Table 2 presents the systems parameters in details.

Table. 2. SIMULATION PARAMETERS

Parameter		Value
Topographical Area (meters)	A	100 m × 100 m
	B	100 m × 100 m
	C	100 m × 100 m
	D	200 m × 50 m
BS or Sink Location (meters)	A	(90, 90)
	B	(90, 90)
	C	(50, 50)
	D	(180, 45)
Number of modes		100
Initial energy of node		0.5 J
Packet data size		2k bit
Eelec		50 nJ/bit
Eamp		100 pJ/bit/m2
Number of transmission packets		2×10^4
Maximum traffic node's queue		10
Limit of transmission distance (meters)		30 m

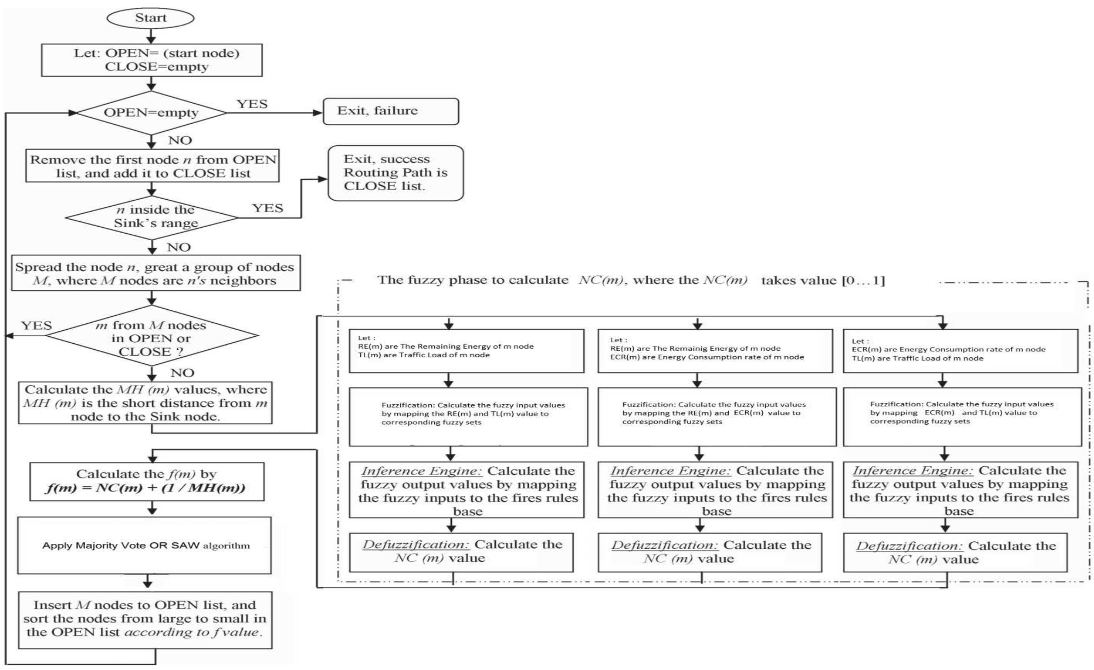


Fig.12.Flow chart of the proposed algorithms

The proposed models assume that the network has the following features:

- Sensor nodes and base station networks are static and immobile.

- Links between nodes are symmetric and approximate distance between each node can be calculated based on the received signal strength (RSSI).
- Flat tracking and routing regardless of make and hierarchical cluster heads is performed.
- All nodes are homogeneous and are identical.
- After calculating the location of every sensor nodes, they are stored within each node.
- All sensor nodes are randomly distributed in the environment.
- All sensor nodes have an equal maximum transmission range and initial energy.
- Each sensor node is waiting to receive a certain amount of traffic in its queue. This queue includes Common applications, traffic that receives to send.
- Channel access methods and data management in wireless sensor network nodes is based on the TDMA model, so that each sensor node send data in its time slot that allocated.

8. METHODS COMPARISION

As Gaikwad method presented in [29] ,the combination of A-Star and two-parameter Fuzzy results in better performance and increase the lifetime of the network ,for this reason we compare it with our proposed techniques ,and use factors such as: number of living nodes, the total energy consumed in each round of data transmission networks, energy consumption at each network data transmission, the residual energy of node and the total amount of packets in the network in order to evaluate and compare at moment of death.

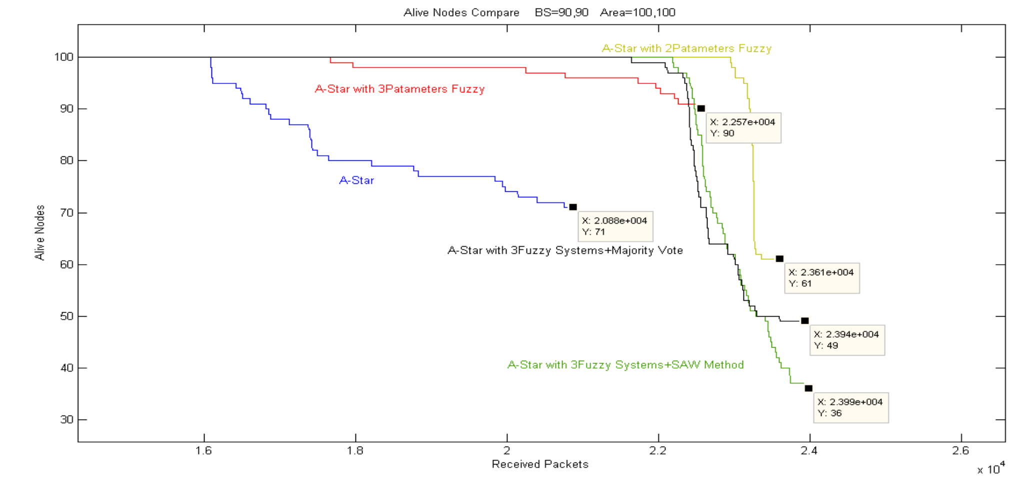
Table .3. The results of the simulation

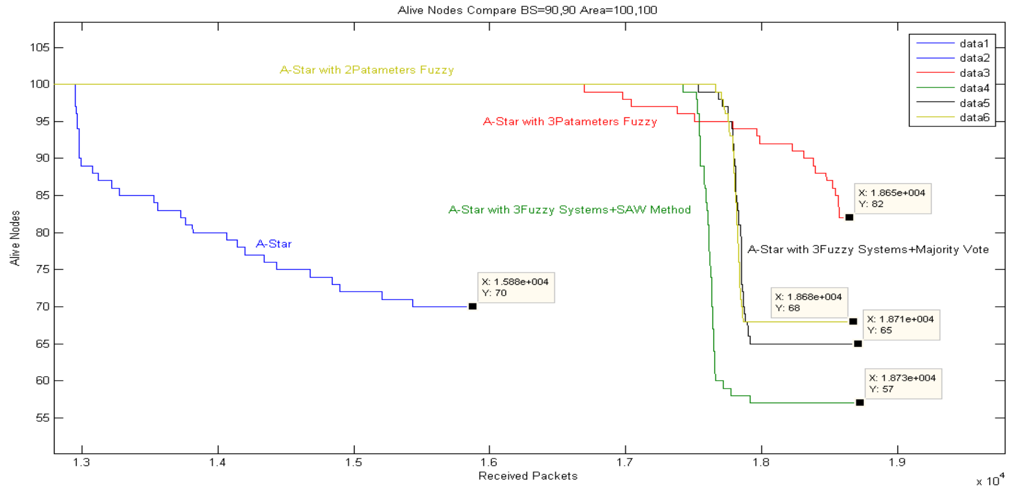
		(1)Area=100x100 , BS=90,90			(2)Area=100x100 , BS=90,90			Area=100x100 , BS=50,50			Area=200x50 , BS=180,45		
No		First Dead Node	Alive Node	Sum Of Packet	First Dead Node	Alive Node	Sum Of Packet	First Dead Node	Alive Node	Sum Of Packet	First Dead Node	Alive Node	Sum Of Packet
1	A-Star	12952	70	15875	16082	71	20907	25555	77	32696	6967	72	9775
2	A-Star 2P Fuzzy	17664	68	18679	22952	61	23605	31584	72	34477	10753	76	11608
3	A-Star 3P Fuzzy	16695	82	18650	17672	90	22571	27513	84	34404	8461	96	9987
4	A-Star 3PF Majority Vote	17537	65	18714	21638	49	23943	32385	51	34495	9523	80	11868
5	A-Star 3PF SAW	17424	57	18730	22191	36	23989	31269	44	34626	10828	81	11914

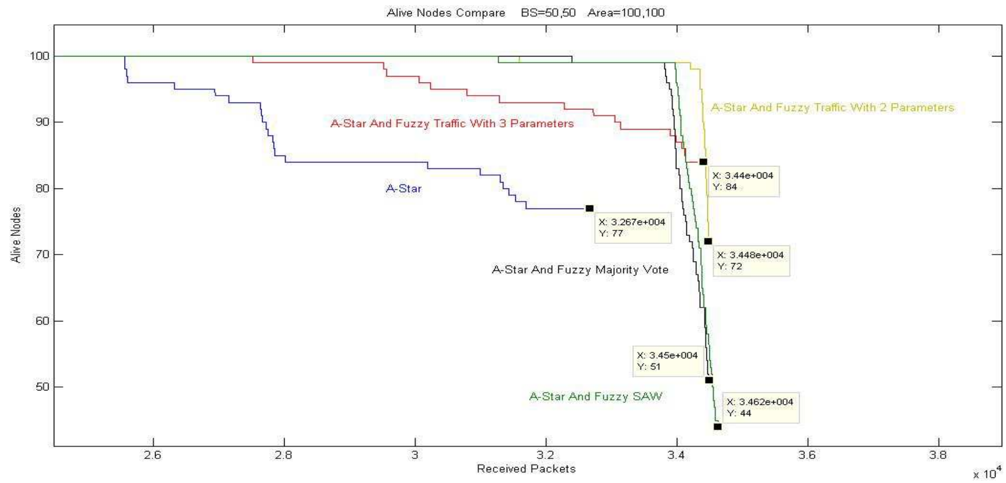
What the chart comparison charts and the rate of improvement and the results of simulations in four topographical environment achieved indicate that in spite of the very small decreasing network lifetime compare to the proposed methods [29], we see a remarkable improvement in the number of nodes alive in A-Star combine by fuzzy with three parameters due to the impact of the rate of energy consumption in routing. A few declines in the network lifetime is due to the efforts of the algorithm in reducing the rate of energy consumption in the nodes with the possibility of choosing a long routes.But at a time, we want to add new BS network in the scenario, due to more nodes alive, A*3PF would be the best way.

Second and third new methods have been presented in this paper that the decision for choosing the best path with regard to energy consumption rate using mix of expert systems with majority of the votes and simple adaptive weight will improve performance and increase network lifetime and this is because using more effective energy consumption rate parameter. Although number of alive nodes a little bit less but this is not the remarkable decline. Between the last two methods of mixing expert systems method, simple adaptive weight (SAW), due to having precise mathematical model ,we see more clear and better improvement. Overall on condition that the aim is themore survival of the network and more data transmission below methods are optimal are the following respectively: A*3PFSAW, A*3PFMV, A*2PF, A*3PF, A*. And in case you want to have more vivid node, the best way to phase three parameters (A*3PF).

https://sites.google.com/site/ijcsis/
ISSN 1947-5500


Alive Nodes Compare BS=90,90 Area=100,100
Alive Nodes
Received Packets
x 10^4
A-Star with 2Patameters Fuzzy
A-Star with 3Patameters Fuzzy
A-Star
A-Star with 3Fuzzy Systems+Majority Vote
A-Star with 3Fuzzy Systems+SAW Method
X: 2.257e+004
Y: 90
X: 2.088e+004
Y: 71
X: 2.361e+004
Y: 61
X: 2.394e+004
Y: 49
X: 2.399e+004
Y: 36


Alive Nodes Compare BS=90,90 Area=100,100
Alive Nodes
Received Packets
x 10^4
data1
data2
data3
data4
data5
data6
A-Star with 2Patameters Fuzzy
A-Star with 3Patameters Fuzzy
A-Star with 3Fuzzy Systems+SAW Method
A-Star
A-Star with 3Fuzzy Systems+Majority Vote
X: 1.865e+004
Y: 82
X: 1.588e+004
Y: 70
X: 1.868e+004
Y: 68
X: 1.871e+004
Y: 65
X: 1.873e+004
Y: 57


Alive Nodes Compare BS=50,50 Area=100,100
Alive Nodes
Received Packets
x 10^4
A-Star And Fuzzy Traffic With 2 Parameters
A-Star And Fuzzy Traffic With 3 Parameters
A-Star
A-Star And Fuzzy Majority Vote
A-Star And Fuzzy SAW
X: 3.44e+004
Y: 84
X: 3.267e+004
Y: 77
X: 3.448e+004
Y: 72
X: 3.45e+004
Y: 51
X: 3.462e+004
Y: 44

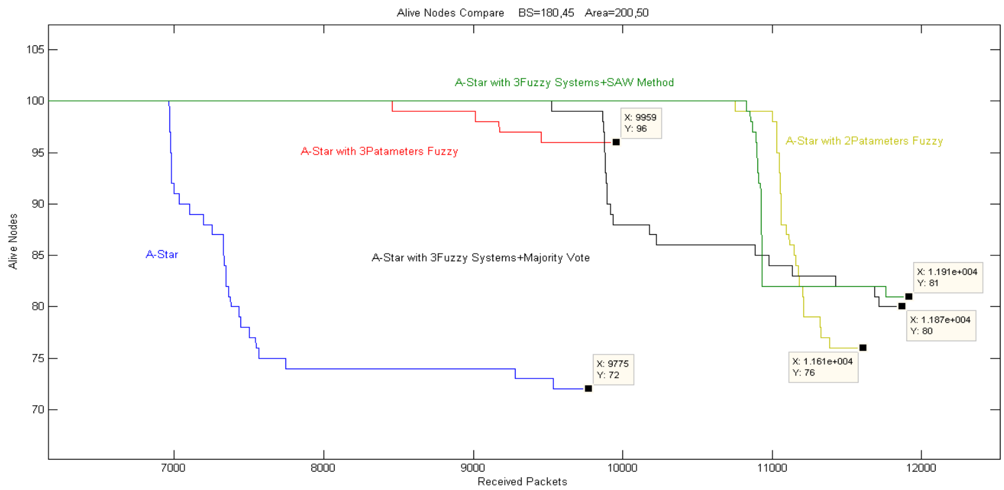


Fig. 13. Comparison chart of the five algorithms in four topographic areas

Table .4. Improvement of the proposed methods in comparison with the A*2PF

	(1)Area=100x100 , BS=90,90						(2)Area=100x100 , BS=90,90					
	First Dead Node	Improvement	Alive Node	Improvement	Sum Of Packet	Improvement	First Dead Node	Improvement	Alive Node	Improvement	Sum Of Packet	Improvement
A-Star 2P Fuzzy	**17664**	-	68	-	18679	-	**22952**	-	61	-	23605	-
A-Star 3P Fuzzy	16695	5.49-	**82**	**20.59**	18650	0.16-	17672	23-	**90**	**47.54**	22571	4.38-
A-Star 3PF Majority Vote	17537	0.72-	65	4.41-	18714	0.19	21638	5.72-	49	19.67-	23943	1.43
A-Star 3PF SAW	17424	1.36-	57	16.17-	**18730**	**0.27**	22191	3.32-	36	40.98-	**23989**	**1.63**

	Area=100x100 , BS=50,50						Area=200x50 , BS=180,45					
	First Dead Node	Improvement	Alive Node	Improvement	Sum Of Packet	Improvement	First Dead Node	Improvement	Alive Node	Improvement	Sum Of Packet	Improvement
A-Star 2P Fuzzy	31584	-	72	-	34477	-	10753	-	76	-	11608	-
A-Star 3P Fuzzy	27513	12.89-	**84**	**16.67**	34404	0.21-	8461	21.31-	**96**	**26.32**	9987	13.96-
A-Star 3PF Majority Vote	32385	2.54	51	29.17-	**34495**	**0.05**	9523	11.44-	**80**	**5.26**	**11868**	**2.24**
A-Star 3PF SAW	31269	1-	44	38.89-	**34626**	**0.43**	10828	0.7	**81**	**6.58**	**11914**	**2.64**

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

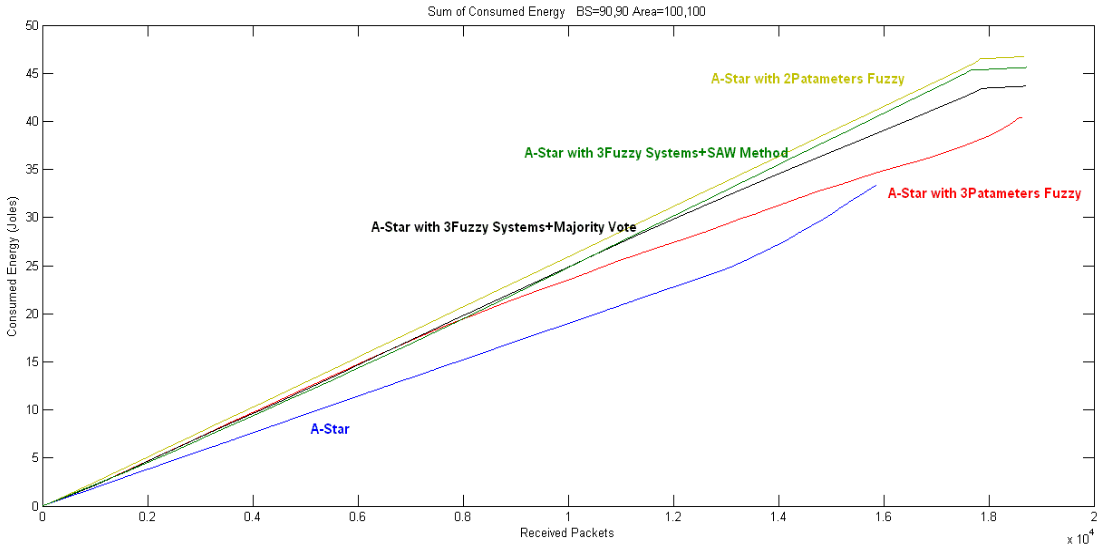


Fig .14. Total network energy consumption in each round of data transmission

Chart above, the total energy consumed in the network is sending data. The lower slope indicates a lower power consumption rate for sending data. So A-Star method initially acts very well, and for sending data energy has low consumption rate, but in the end, it has high slope in energy consumption and will be more than the other methods, which causes more and rapid consumption of energy and therefore nodes in the network will die sooner. Fuzzy method with two parameters has the highest rate of energy consumption. Since the rate of energy consumption is not involved in the selection of the route, so it's pretty uniform and totally linear in power consumption curve.But in the Majority Vote and SAW method with balanced energy consumption, we witness increasing in network lifetime and more data sent. In A*3PF approach the more pass network lifetime, the more changing in energy consumption rate and this approach adapt itself to the conditions ,so the number of alive nodes in the end will be more.

The below graphs represent the amount of energy nodes in the network death time. it shows that network in the A* method will die while many nodes are still high energy and alive In A*2PF approach many nodes die while than next method less packets sent , but in A*3PF , A*3PFSAWand A*3PFMV approaches the situation is better ,the nodes with high energy vs. the nodes with low energy is almost balanced.The balances of energy consumption of all nodes in the network have been done better. The Majority Vote and SAW methods regardless of the increasing network lifetime, we will see balance in the energy consumption at each round ,which indicates that these two methods is better.

To sum up of comparisons it can be said that the death of the first node and the number of nodes remaining alive in the last round and the remaining amount of energy networks in the different algorithms are affected by some factors like: the movement of the geographical location of the BS and change the size of the network environment and make different network behavior.

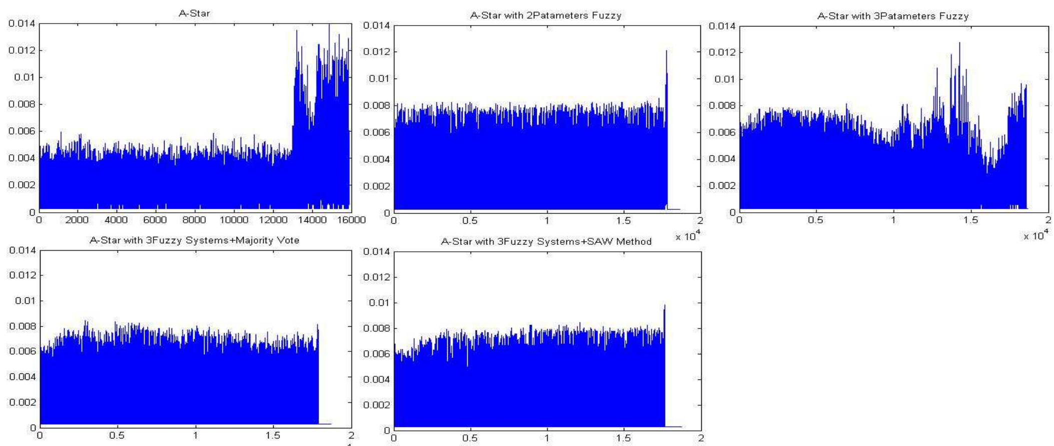


Fig. 15. Amount of network energy used in each time of data transmission (round)

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

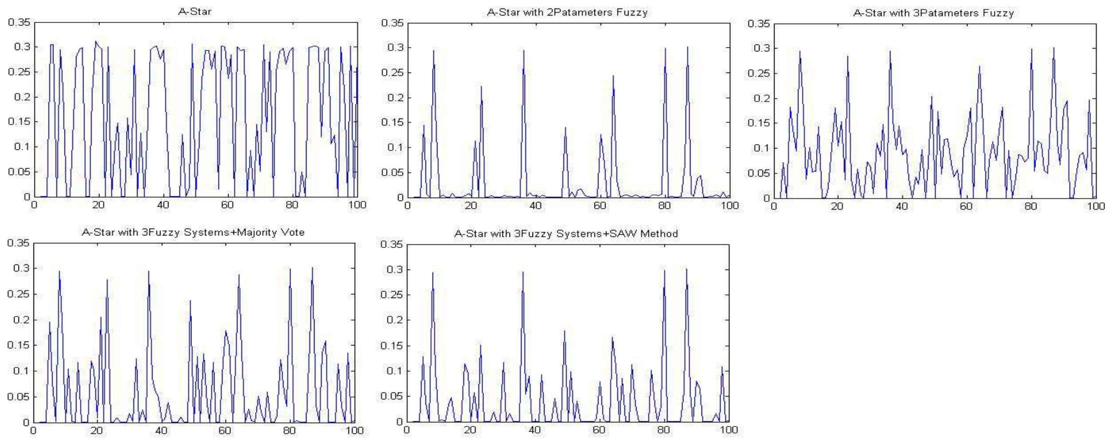


Fig. 16. Energy levels of network nodes at the moment of death.

Table .5. Summary of the proposed methods in comparison

<table>
<tr><td></td><td colspan="4">First Dead Node</td></tr>
<tr><td>Rank</td><td>Topography Area 100 100 - BS 90 90</td><td>Topography Area 100 100 - BS 90 90-New Initial</td><td>Topography Area 100 100 - BS 50 50</td><td>Topography Area 200 50 - BS 180 45</td></tr>
<tr><td>1</td><td colspan="2">A-Star with 2Patameters Fuzzy</td><td>A-Star with 3Fuzzy Systems+Majority Vote</td><td>A-Star with 3Fuzzy Systems+SAW Method</td></tr>
<tr><td>2</td><td>A-Star with 3Fuzzy Systems+Majority Vote</td><td>A-Star with 3Fuzzy Systems+SAW Method</td><td colspan="2">A-Star with 2Patameters Fuzzy</td></tr>
<tr><td>3</td><td>A-Star with 3Fuzzy Systems+SAW Method</td><td>A-Star with 3Fuzzy Systems+Majority Vote</td><td>A-Star with 3Fuzzy Systems+SAW Method</td><td>A-Star with 3Fuzzy Systems+Majority Vote</td></tr>
<tr><td>4</td><td colspan="4">A-Star with 3Patameters Fuzzy</td></tr>
<tr><td>5</td><td colspan="4">A-Star</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td colspan="4">Alive Nodes on Last state</td></tr>
<tr><td>Rank</td><td>Topography Area 100 100 - BS 90 90</td><td>Topography Area 100 100 - BS 90 90-New Initial</td><td>Topography Area 100 100 - BS 50 50</td><td>Topography Area 200 50 - BS 180 45</td></tr>
<tr><td>1</td><td colspan="4">A-Star with 3Patameters Fuzzy</td></tr>
<tr><td>2</td><td colspan="3">A-Star</td><td>A-Star with 3Fuzzy Systems+SAW Method</td></tr>
<tr><td>3</td><td colspan="3">A-Star with 2Patameters Fuzzy</td><td>A-Star with 3Fuzzy Systems+Majority Vote</td></tr>
<tr><td>4</td><td colspan="3">A-Star with 3Fuzzy Systems+Majority Vote</td><td>A-Star with 2Patameters Fuzzy</td></tr>
<tr><td>5</td><td colspan="3">A-Star with 3Fuzzy Systems+SAW Method</td><td>A-Star</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td colspan="4">Sum of Packet Sent (Life Time Of The Network)</td></tr>
<tr><td>Rank</td><td>Topography Area 100 100 - BS 90 90</td><td>Topography Area 100 100 - BS 90 90-New Initial</td><td>Topography Area 100 100 - BS 50 50</td><td>Topography Area 200 50 - BS 180 45</td></tr>
<tr><td>1</td><td colspan="4">A-Star with 3Fuzzy Systems+SAW Method</td></tr>
<tr><td>2</td><td colspan="4">A-Star with 3Fuzzy Systems+Majority Vote</td></tr>
<tr><td>3</td><td colspan="4">A-Star with 2Patameters Fuzzy</td></tr>
<tr><td>4</td><td colspan="4">A-Star with 3Patameters Fuzzy</td></tr>
<tr><td>5</td><td colspan="4">A-Star</td></tr>
</table>

9. CONCLUSION

In wireless sensor networks where nodes operate on limited battery energy efficient utilization of the energy is very important. One of the main characteristics of these networks is that the network lifetime is highly related to the route selection. Unbalanced energy consumption is an inherent problem in a WSN. To efficiently route data through transmission path from node to node and to prolong the overall lifetime of the network, we proposed some new algorithms by using a combination of both Mix-Fuzzy approach and A-star algorithm. The new method is capable of selecting optimal routing path from the source node to the sink by favoring the highest remaining energy, minimum number of hops, lowest traffic load and lowest energy consumption rate. The performance of the proposed method is evaluated and compared with other methods under the same criteria in four different topographical areas. by using mix of expert system and proper use of three parameters: remaining energy, traffic load and energy consumption rate in their fuzzy system Simulation results demonstrate the effectiveness of the new approaches -A*3PFSAW and A*3PFMV- than A*2PF, A*3PF, A* methodswith regards to enhancement of the lifetime of wireless sensor networks with randomly scattered nodes.

However according the results and collected statistics should be noted that the network performance improvement in terms of the maximizing lifetime will be affected by such as geographical placement BS and the size (length and width) of the network, the number of sensor nodes, neighborhood radius, BS and node mobility, the amount of initial network energy, heuristic algorithm, the type of algorithm used for mixing experts systems ,nodes distribution in the network environment, node

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

density, etc. but to sum up, our new proposed algorithms - A*3PFSAW, A*3PFMV - in the improvement of performance and network lifetime for different scenarios implemented have better performance and stability of the algorithm in such a network.

10. References

1- R.V.Kulkarni,A.Forster,andG. K.Venayagamoorth, "Computational intelligence inwirelesssensornetworks:A survey,"*IEEECommun. SurveysTutorials*,vol.13,no.1,pp.68–96,Feb.2011.

2- C. Hua and T. P. Yum, "Optimal routing and data aggregation for maximizing lifetime of wireless sensor networks," IEEE ACM Trans. Netw., vol. 16, no. 4, pp. 892–903, Aug. 2008.

3- H. Zhang and H. Shen, "Balancing energy consumption to maximize network lifetime in data-gathering sensor networks," IEEE Trans. Par-allelDistrib. Syst., vol. 20, no. 10, pp. 1526–1539, Oct. 2009.

4- J. N. Al-Karaki and A. E. Kamal, "Routing techniques in wireless sensor networks: A survey," IEEE Wireless Commun., vol. 11, no. 6, pp. 6–28, Dec. 2004.

5- H. R. Karkvandi, E. Pecht, and O. Yadid, "Effective lifetime-aware routing in wireless sensor networks," IEEE Sensors J., vol. 11, no. 12, pp. 3359–3367, Dec. 2011.

6- K. Akkaya and M. Younis, "A survey of routing protocols in wireless sensor networks," Ad Hoc Netw., vol. 3, no. 3, pp. 325–349, May 2005.

7- F. Ren, J. Zhang, T. He, C. Lin, and S. K. Das, "EBRP: Energy-balanced routing protocol for data gathering in wireless sensor networks," IEEE Trans. Parallel Distrib. Syst., vol. 22, no. 12, pp. 2108–2125, Dec. 2011.

8- M. J. Tsai, H. Y. Yang, and W. Q. Huang, "Axis-based virtual coordinate assignment protocol and delivery-guaranteed routing protocol in wireless sensor networks," in Proc. IEEE INFOCOM 26th Int. Conf. Comput. Commun., May 2007, pp. 2234–2242.

9- J. Park and S. Sahni, "An online heuristic for maximum lifetime routing in wireless sensor networks," IEEE Trans. Comput., vol. 55, no. 8, pp.1048–1056, Aug. 2006.

10- C. Wu, R. Yuan, and H. Zhou, "A novel load balanced and lifetime maximization routing protocol in wireless sensor networks," in Proc. IEEE Vehicular Technol. Conf. VTC Spring, May 2008, pp. 113–117.

11- C. C. Hung, K. C. J. Lin, C.-C. Hsu, C.-F. Chou, and C.-J. Tu, "On enhancing network-lifetime using opportunistic routing in wireless sensor networks," in Proc. 19th Int. Conf. Comput. Commun. Netw.,Aug. 2010, pp. 1–6.

12- R. Madan and S. Lall, "Distributed algorithms for maximum lifetime routing in wireless sensor networks," IEEE Trans. Wireless Commun., vol. 5, no. 8, pp. 2185–2193, Aug. 2006.

13- J. H. Chang and L. Tassiulas, "Maximum lifetime routing in wireless sensor networks," IEEE/ACM Trans. Netw., vol. 12, no. 4, pp. 609–619, Aug. 2004.

14- O. Zytoune, M. El-Aroussi, and D. Aboutajdine, "A uniform balancing energy routing protocol for wireless sensor networks," Wireless Personal Commun., vol. 55, no. 2, pp. 147–161, Oct. 2010.

15- Y. M. Lu and V. W. S. Wong, "An energy-efficient multipath routing protocol for wireless sensor networks," in Proc. IEEE 64th Vehicular Technol. Conf., Sep. 2006, pp. 1–5.

16- C. Park and I. Jung, "Traffic-aware routing protocol for wireless sensor networks," in Proc. IEEE Inform. Sci. Appl. Int. Conf., Apr. 2010, pp.1–8.

17- M. R. Minhas, S. Gopalakrishnan, and V. C. M. Leung, "An online multipath routing algorithm for maximizing lifetime in wireless sensor networks," in Proc. IEEE Inform. Technol. New Generat. 6th Int. Conf., Apr. 2009, pp. 581–586.

18- M. A. Azim and A. Jamalipour, "Performance evaluation of optimizedforwarding strategy for flat sensor networks," in Proc. IEEE GlobalTelecommun. Conf., Nov. 2007, pp. 710–714.

19- S. Y. Chiang and J. L. Wang, "Routing analysis using fuzzy logicsystems in wireless sensor networks," Lecture Notes Comput. Sci., vol.5178, pp. 966–973, 2008.

20- K. M. Rana and M. A. Zaveri, "ASEER: A routing method to extend life of two-tiered wireless sensor network," Int. J. Adv. Smart Sensor Netw. Syst., vol. 11, no. 2, pp. 1–16, Oct. 2011.

21- W. Dargie and C. Poellabauer, "Network layer," in Fundamental of Wireless Sensor Networks Theory and Practice. New York: Wiley, 2010, pp. 163–204.

22- L. A. Zadeh, "Soft computing and fuzzy logic," IEEE Software, vol. 11, no. 6, pp. 48–56, Nov. 1994.

23- K.-Y. Cai and L. Zhang, "Fuzzy reasoning as a control problem," IEEETrans. Fuzzy Syst., vol. 16, no. 3, pp. 600–614, Jun. 2008.

24- T. A. Runkler, "Selection of appropriate defuzzification methods using application specific properties," IEEE Trans. Fuzzy Syst., vol. 5, no. 1, pp. 72–79, Feb. 1997.

25- K. M. Passino and P. J. Antsaklis, "A metric space approach to the specification of the heuristic function for the A* algorithm," IEEE Trans. Syst. Man Cybern., vol. 24, no. 1, pp. 159–166, Jan. 1994.

26- X. H. Li, S. H. Hong, and K. L. Fang, "WSNHA-GAHR: A greedy and A* heuristic routing algorithm for wireless sensor networks in home automation," IET Commun., vol. 5, no. 13, pp. 1797–1805, Sep. 2011.

27- J. Yao, C. Lin, X. Xie, A. J. Wang, and C. C. Hung, "Path planning forvirtual human motion using improved A* star algorithm," in Proc. IEEE Inform. Technol. New Generat. 7th Int. Conf., Apr. 2010, pp. 1154–1158.

28- W. Heinzelman, A. Chandrakasan and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Mi- crosensor Networks," Proceedings of the 33rd Hawaii International Conference on System Sciences (HICSS '00), January 2000.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

29- Deepak S. Gaikwad, SampadaPimpale ,Routing Alternatives for Network Lifetime Maximization of WSNs Using Heuristic and Fuzzy Logic Approach , International Journal of Inventive Engineering and Sciences (IJIES) ISSN: 2319–9598, Volume-1, Issue-6, May 2013

11. APPENDIX

Table A: Fuzzy Rules

FIS 1		
Remain Energy	Traffic Load	Output
VL	VL	L
VL	L	VL
VL	M	VL
VL	H	VL
VL	VH	VL
L	VL	M
L	L	M
L	M	L
L	H	L
L	VH	VL
M	VL	H
M	L	M
M	M	M
M	H	L
M	VH	L
H	VL	VH
H	L	H
H	M	H
H	H	M
H	VH	M
VH	VL	VH
VH	L	VH
VH	M	VH
VH	H	H
VH	VH	H

FIS 2		
Remain Energy	Energy Consumtion Rate	Output
VL	VL	L
VL	L	L
VL	M	VL
VL	H	VL
VL	VH	VL
L	VL	M
L	L	M
L	M	L
L	H	L
L	VH	VL
M	VL	H
M	L	M
M	M	M
M	H	L
M	VH	L
H	VL	VH
H	L	H
H	M	H
H	H	M
H	VH	M
VH	VL	VH
VH	L	VH
VH	M	VH
VH	H	H
VH	VH	M

FIS 3		
Traffic Load	Energy Consumtion Rate	Output
VL	VL	VH
VL	L	H
VL	M	M
VL	H	L
VL	VH	L
L	VL	H
L	L	M
L	M	M
L	H	L
L	VH	L
M	VL	H
M	L	H
M	M	M
M	H	L
M	VH	L
H	VL	M
H	L	M
H	M	L
H	H	L
H	VH	VL
VH	VL	L
VH	L	L
VH	M	L
VH	H	VL
VH	VH	VL

FIS			
Traffic Load	Remain Energy	Energy Consumtion Rate	Output
VL	VL	VL	L
VL	VL	L	L
VL	VL	M	L
VL	VL	H	VL
VL	VL	VH	VL
VL	L	VL	L
VL	L	L	L
VL	L	M	L
VL	L	H	VL
VL	L	VH	VL
VL	M	VL	M
VL	M	L	M
VL	M	M	L
VL	M	H	L
VL	M	VH	L
VL	H	VL	H
VL	H	L	H
VL	H	M	M
VL	H	H	M
VL	H	VH	L
VL	VH	VL	VH
VL	VH	L	H
VL	VH	M	H
VL	VH	H	M
VL	VH	VH	M
L	VL	VL	L
L	VL	L	L
L	VL	M	L
L	VL	H	VL
L	VL	VH	VL
L	L	VL	L
L	L	L	L
L	L	M	L
L	L	H	VL
L	L	VH	VL
L	M	VL	M
L	M	L	L
L	M	M	L
L	M	H	L
L	M	VH	L
L	H	VL	H
L	H	L	H
L	H	M	M
L	H	H	L
L	H	VH	L
L	VH	VL	VH
L	VH	L	H
L	VH	M	H
L	VH	H	M

Allocation Algorithm based on CAC Scheme for LTE Network

Radhia Khdhir
LETI Laboratory
ENIS, University of Sfax, Tunisia

Kais Mnif
LETI Laboratory
ENETCOM, University of Sfax, Tunisia

Khitem Ben Ali
LETI Laboratory
ENIS, University of Sfax, Tunisia

Lotfi Kammoun
LETI Laboratory
ENIS, University of Sfax, Tunisia

Abstract—To reduce network congestion and to guarantee a certain level of Quality of Service (QoS) for service requests, Call Admission Control (CAC) as a part of Radio Resource Management (RRM) aims to accept or reject a call based on available resources. In this paper, we proposed new CAC and resources allocation schemes for Long Term Evolution (LTE). The proposed CAC scheme gives the priority of Handoff Calls (HC), without **totally** neglecting the requirements of a New Calls (NC). The main objective of this approach is to provide QoS and to prevent network congestion. Simulation results show that the call admission control scheme leads to increased session establishment success and resource utilization compared with existing admission control and resources allocation schemes. Moreover, the resources allocation scheme achieves a considerable gain in the system throughput and fairness.

Keywords—Call admission control; QoS; Scheduling; LTE; Uplink; Throughput.

I. INTRODUCTION

The Orthogonal Frequency Division Multiple Access (OFDMA) and Single Carrier Frequency Division Multiple Accesses (SC-FDMA) are the respective techniques used for radio transmission and reception in LTE and Long Term Evolution Advanced (LTE-A) networks for the Downlink (DL) and Uplink (UL) directions, respectively. SC-FDMA offers improvements in terms of spectral efficiency and throughput while satisfying several types of services. Indeed, The LTE and LTE-A systems are expected to provide peak data rates in the order of 50 and 500 Mbit/s in uplink, respectively [1]. In the LTE and LTE-A uplink directions, the total bandwidth is divided into multiple sub-bandwidths. These sub-bandwidths are regrouped in Physical Resource Blocks (PRBs). A PRB is defined by a couple of frequency and time domains. In fact, a PRB is 0.5 ms in length (one slot in the time domain) and contains a contiguous set of 12 subcarriers (180 kHz in the frequency domain) for each OFDM symbol. Therefore, this PRB is the basic transmission unit of a user's data in both uplink and downlink directions. In order to provide quality of services (QoS) for different kinds of services in packet switched networks, RRM can be of a great importance.

LTE standards do not specify any Call Admission Control (CAC) and resources allocation algorithms have to be defined and so are left to the vendors and the researchers to implement them [2],[3]. The CAC decides whether the eNodeB accepts or rejects the call requests of User Equipments (UEs) by considering the cell capacity. The scheduler, then, selects the accepted requests to be scheduled in the following Transmission Time Interval (TTI) based on their QoS requirements. For the allocation scheme, the eNodeB needs some channel quality information perceived by each UE. This is achieved by sending Sounding Reference Signal (SRS) from UEs to the eNodeB so that the latter can compute the Channel Quality Indicator (CQI) values of each PRB for each UE.

In this paper, we propose a new CAC scheme that handles the HC and NC and increases session establishment success and resource utilization. Then, we present a new scheduler that treats both Guaranteed Bit Rate (GBR) and Non Guaranteed Bit Rate (NGBR) traffics, by taking the maximization throughput and the user fairness into consideration.

The rest of this paper is organized as follows: section II presented pre-studied CAC and resources allocation algorithms. The system model, proposed CAC and resources allocation algorithms were introduced in section III. The simulation results and discussions were detailed in section IV. Finally, we drew our conclusions in section V.

II. LITERATURE REVIEW

We can summarize the existing CAC and scheduling algorithms by the following conclusions

-CAC schemes treat all calls equally (HC and NC) and do not differentiate them relying on their type. This is the case of proposed CAC schemes in [4], [6],[7],[8] and [9].

-CAC schemes prioritize the HC over NC. So, they neglect the NC. This is the case of CAC schemes elaborated in [10], [11], [12] and [13].

- Schedulers do not consider QoS requirements of different applications and multiclass traffics. So, they handle the GBR and NGBR traffics with same principle. This is the case of

schedulers elaborated in [14],[16],[19],[20],[21],[22] and [23].
- Schedulers do not consider fairness among users. This is the case of schedulers elaborated in [15], [19] and [24].
Hence, there is a need for a CAC scheme that supports both HC and NC and increases session establishment success **and resource utilization**. To tackle these objectives, we design a new CAC scheme. Mainly, we use HC and NC queues. Then, we attribute the high priority for primary queue without neglecting the NC. Indeed, we adjust the threshold, according to the network conditions, to guarantee that sufficiently resources will be available for the HC. Finally, transmissions will be performed based on our proposed scheduler named Robust Uplink Packet Scheduling Algorithm (RUPSA). This scheduler handles both GBR and NGBR traffics. The principle of our proposal as well as its performance analysis will be discussed in the next sections.

III. SYSTEM MODEL AND PROPOSED SCHEDULING ALGORITHMS

III.1 SYSTEM MODEL

We consider the Evolved Packet System (EPS) with one eNodeB (is the entity in charge of performing the resource allocation), *m* PRBs and *n* active UEs. The EPS bearers are classified into two types: GBR and NGBR. The objective of the CAC functionality is to determine whether a new EPS bearer can be activated (CAC is responsible of accepting or rejecting a connection depending on network available resources). In our system model, the number of users is 120 and their position is uniformly distribution at the starting of simulation. The random-walk model is considered as the mobility model. Requests arrive at eNodeB as Poisson processes with parameter λ. Then service time is measured by an exponential distribute with mean 1/μ.

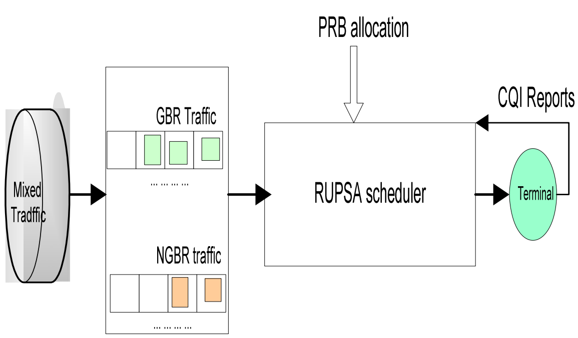


Figure.1.RUPSA illustration

The packets coming to the network from mixed traffic are classified into two queues GBR and NGBR classes. Then, each class of packet will be delivered in independent queue. These two queues will be served on the basis of RUPSA. An illustration of the proposed CAC and uplink scheduling transmissions is shown in Figure 1.

In our proposed algorithm RUPSA, we introduce a weighting factor ρ which represents the portion of the reserved resources blocks for GBR users for the total available PRBs. By using the weighting factor, we guarantee that sufficiently resources will be available for the GBR users.

III.2 PROPOSED SCHEMES

In this section, we present a new CAC and scheduling algorithms for an LTE system.

III.2.1 CAC Scheme

In this subsection we propose a CAC scheme for the LTE network, which provides a PRBs allocation policy that takes into account the distinction between incoming traffic for each class and prioritizes HC, without neglecting **NC**.
The objective of the CAC scheme is to improve resource utilization and decrease the dropping probability. The input of our CAC scheme is the following QoS parameters:
$D_{req(k)}, D_{max_{GBR}}, D_{max_{NGBR}}, RB_{req}, RB_{min}, RB_{avail}, RB_{reser_{GBR}}, Pr, lengHC$ and ρHC.
Where:
$D_{req(k)}$: The delay of user request k

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

$D_{max_{GBR}}$: The delay budget which is the upper delay bound of GBR traffic
$D_{max_{NGBR}}$: The delay budget which is the upper delay bound of NGBR traffic
RB_{req}: The required number of PRBs
RB_{min}: The minimum number of required PRB
RB_{avail} : The number of available PRBs
$RB_{reserGBR}$: The number of reserved PRBs for the GBR traffic
Pr: The type of call, HC or NC
$lengHC$: The length of HC queue
ρHC: The threshold size of the HC queue

When the call arrives to the network; the eNodeB is capable to identify its type at any time *t* based on the receiving QoS parameters. In our work, we provide a CAC scheme that takes into account the distinction between incoming traffic for each class and prioritizes HC over NC, without neglecting NC.
Then, we assign two service classes for the coming calls (GBR and NGBR traffic) depending on their QoS parameters. The algorithm proposes a system of priority for the four service classes in the increasing direction: NC-NGBR, HC-NGBR, NC-GBR and HC-GBR. The calls coming in mixed traffic in similar types (HC or NC) to an overloaded cell will be classified into specific queues (HC queue and NC queue). Since, the latency of these calls depends on the type of traffic; the calls will be handled differently. In the ideal case, all calls in a cell should be allocated RB_{req} whenever possible. However, in overloaded cell, some of the calls receive a lower bandwidth than requested.
For the NC buffered in the NC queue, initially, the "**lengHC<ρHC"** condition must be checked to satisfy the HC prioritization over the NC. The flow chart of our proposed scheme is shown in Figure 2.

The CAC algorithm steps are as follows:

Step 1: Calls arrive specifying their QoS parameters like $\boldsymbol{D_{req(k)}}, \boldsymbol{D_{max_{GBR}}}, \boldsymbol{D_{max_{NGBR}}}, \boldsymbol{RB_{req}}, \boldsymbol{RB_{min}}$ $RB_{avail}, RB_{reser_{GBR}}, Pr$, lengHC and ρHC.

Step 2: The call type (NC or HC) is determined.

Step 3: **(a)** If the number of PRBs is sufficient then the call is accepted.
(b) Else
(i) If this call is NC, the condition $\boldsymbol{leng_{HC}}<\boldsymbol{\rho}_{HC}$ is checked. If true then proceed to next step, else the call is rejected.
(ii) If this call is HC, then proceed to next step.

Step 4: LTE call type (GBR or NGBR) is determined.

Step5: The condition on the latency delay is checked ($D_{req(k)}<D_{max_{GBR}}$ if the call is GBR type or $D_{req(k)}<D_{max_{NGBR}}$ if it is an NGBR call), if true then proceed to next step, else the call is rejected.

Step6: The condition on the sufficiency PRBs is checked (if this call is NC-GBR type then $RB_{req(k)}<RB_{reserGBR}$ is checked, else if this call is HC-GBR $RB_{req(k)}<RB_{avail}$ is checked, else if this call is NC-NGBR or HC-NGBR then $RB_{req(k)}<(RB_{avail}-RB_{reserGBR})$ is checked
(a) For the NCs, if no resources are available the call is rejected, else the call is accepted.
(b) For the HCs, proceed to next step

Step7: The condition on the sufficiency of PRBs versus RB_{min} is checked ($RB_{min}<RB_{avail}$ is checked for HC-GBR calls and $RB_{min}<(RB_{avail}-RB_{reserGBR})$ is checked for HC-NGBR calls.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

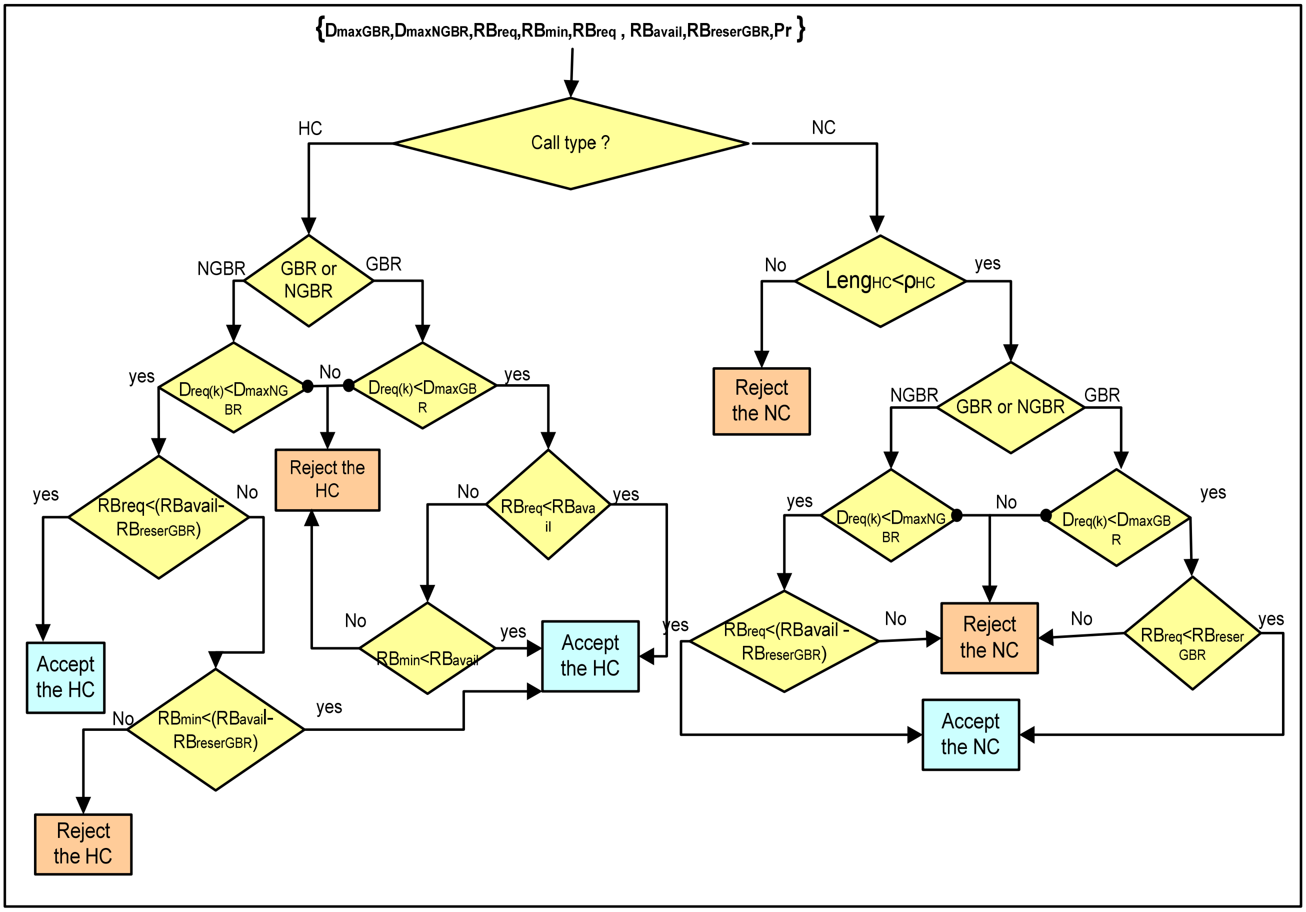


Figure.2. Flow Chart of the proposed CAC algorithm

III.2.2 Scheduling Scheme

In this subsection, we present our proposed algorithm (RUPSA) for an Uplink LTE system. RUPSA serves GBR and NGBR packets, classified into two independent queues, using the proposed priorities function (7). The proposed priorities function handles two principal objectives: throughput and fairness.

RUPSA aims to maximize the throughput. So, the first optimization problem can be mathematically defined as follows:

$$Max \sum_{i=1}^{n} \omega_i R_i \quad (1)$$

$$\mathbb{C}_i \cap \mathbb{C}_j = \emptyset, \quad \forall\, i \neq j,\ i \in I \text{ and } j \in I \quad (2)$$

$$\mathbb{C}_1 \cup \mathbb{C}_2 \ldots \cup \mathbb{C}_n \subseteq \mathbb{C} \quad (3)$$

Where R_i is the average throughput for user i,ω_i is the QoS weight for user i,$\mathbb{C}$ is the set of available PRBs, $\mathbb{C}_i$ is the set of PRBs assigned to user i, n is the total number of users and I is the set of users. The constraint of this algorithm is to assign each PRB to only one user i without any overlap.

We define the weighting factor ω_i as follows:

$$\boldsymbol{\omega_i} = \begin{cases} \boldsymbol{\rho} & \boldsymbol{for\ GBR\ traffic} \\ \boldsymbol{1-\rho} & \boldsymbol{for\ NGBR\ traffic} \end{cases} \quad (4)$$

Where ρ represents the portion of the reserved resource blocks for GBR traffic among available PRBs.

In addition to the throughput maximization, our second objective is to guarantee fairness basing on the fairness scheduling method proposed in [22] (see equation 5).

$$\mathrm{F}_{\boldsymbol{i}} = \frac{\boldsymbol{R_i}}{GBR_i} \quad \textbf{(5)}$$

Where F_i the capability weight is calculated at each TTI and GBR_i is the guaranteed bit rate of the user's application or service flow. In this work, we modified equation (5) differently as follows:

$$\mathrm{F}_{\boldsymbol{i}} = \boldsymbol{R_i} / \boldsymbol{\Re_{i(req)}} \quad (6)$$

Where $\Re_{i(req)}$ Represents the minimum required throughput. Two cases are considered:

- $\Re_{i(req)}$ represents the guaranteed bit rate for GBR users [26]
- $\Re_{i(req)}$ Represents the minimum throughput that would be considered acceptable for NGBR users.

Let R_i^{alloc} be the number of bits that can be transmitted in a subframe for user i. As a result, during a subframe s, the

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

eNodeB should try to allocate PRBs in a way that allows R_i^{alloc} bits to be transmitted on average.The number of bits that can be transmitted by the allocated PRBs depends on the corresponding CQI values as shown in Table III.
Before allocating PRBs, the eNodeB has to decide about the priority of each user *i*. So, in each TTI, the user with the highest priority metric, using equation (7), is selected to schedule. We define the priority metric as follows:

$$\boldsymbol{P_i}(\boldsymbol{s}) = \mathrm{F}_{\boldsymbol{i}} \times \boldsymbol{\omega_i} \times \frac{\boldsymbol{R_i^{alloc}} - \boldsymbol{R_i}(\boldsymbol{s}-\mathbf{1})}{\boldsymbol{R_i^{alloc}}} \qquad (7)$$

The equation (7) represents the function priority calculated each TTI. By this function two objectives are handled: fairness is presented by F_{i} and the throughput is presented by $\frac{\mathrm{R_i^{alloc}} - \mathrm{R_i}(s-1)}{\mathrm{R_i^{alloc}}}$.

Table I.THE CQI PARAMETRS[22]

CQI index	Modulation	Code rate *1024	efficiency	Bits per PRB per subframe
1	QPSK	78	0.1523	21.931
2	QPSK	120	0.2344	33.754
3	QPSK	193	0.3770	54.288
4	QPSK	308	0.6016	86.630
5	QPSK	449	0.8770	126.288
6	QPSK	602	1.1758	169.315
7	64QAM	378	1.4766	212.630
8	64QAM	490	1.9141	275.630
9	64QAM	616	2.4063	346,507
10	64QAM	466	2.7305	393.192
11	64QAM	567	3.3223	478.411
12	64QAM	666	3.9023	561.931
13	64QAM	772	4.5234	651.370
14	64QAM	873	5.1152	736.589
15	64QAM	948	5.5547	799.877

The steps of the proposed scheduling algorithm are as follows:

Step 1: Initialize the set $\mathbb{C}$ of the available PRBs for allocation.

Step 2: Calculate the priority of users set *I* based on equation **(7)**

Step 3: Select the user *i* with the highest priority calculated by equation **(7).**

Step 4: Assign the PRB with the highest CQI value to selected user *i*,

(a) If the number of bits that can be transmitted by the allocated PRBs is smaller than the number of bits granted by the minimum required throughput ($\mathfrak{R}_{\mathrm{i(req)}}$), then search and include free adjacent PRBs on both sides to increase the number of bits until the number of required bits is achieved.

(b) Otherwise, cancel this allocation and search the PRB corresponding of the second highest CQI value. Allocate this PRB, to the selected user (step3).

Step 5: Remove the set of PRBs allocated to user *i* from the $\mathbb{C}_i$:$\mathbb{C} = \mathbb{C} - \mathbb{C}_{\boldsymbol{i}}$

Step6: Remove the user *i* from set *I : I=I-i*

Step7: Repeat the steps from 2 to 6 until all PRBs are allocated or all users are served.

The complexity analysis of scheduling algorithms is based on the number of iterations an algorithm achieves when searching for the final allocation (user-PRB). RUPSA, allocates each PRB after completing a linear search on the PRBs and UEs in order to find the UE-PRB pair that maximizes the priority value (equation 7). Consequently, the complexity of the algorithm is ***O(nm)***. Recall that *n* is the total number of users and *m* is the total number of PRBs.

IV. RESULTS AND DISCUSSION

In this section we present the simulation results obtained by applying the proposed algorithms in section *III.*

IV.1 Simulation Parameters

In order to study the performance of the proposed CAC scheme, we use the standard generated in 3GPP deployment evaluation parameters [27]. More details on the configuration parameters used in this simulation are given in Table IV.

Table II.SIMULATION PARAMETERS

Parameters	Value
System bandwidth	20 MHz
Subcarrier spacing	15 KHz
Number of subcarriers per PRB	12
Number of available PRB	100
Transmission time interval(TTI)	1 ms
Total number of used subcarriers	1200
Carrier frequency	2.5 GHz
Frame duration	10 ms
Slot duration	0.5 ms
Number of users	50
ρ	0.7
Simulation Time	1000 TTIs
Link adaptation ACM Modulation	BPSK, QPSK,16-QAM, 64-QAM
Scheduling algorithms	RR, AAG-R, RME and RUPSA

IV.2 Simulation Results

In this subsection, we evaluate the performance of our proposed schemes in terms of HC dropping probability, NC blocking probability, served users, system throughput, end to end delay and fairness.

IV.2.1 Handoff call dropping/New call blocking probability

HC dropping probability (HCDP) is defined as the fraction of handoff attempts that are denied access because of lack of resources. NC blocking probability (NCBP) is defined as the fraction of NCs that are blocked because of lack of resources.

In Figures 3 and 4, we can see that if we increase the number of UEs, it leads to increase in HCDP and in NCBP. This is because the increase in the number of occupied PRBs causes the loading of the network.

Figure 3 shows the HCDP and NCBP for GBR traffic of our proposed scheme and that proposed in [10]. It is clear that if we apply the proposed CAC scheme a decrease in the blocking rate is guaranteed compared to the solution proposed in [10]. The growth starts from a number of users equal to 20 for our proposed CAC algorithm.

When applying our scheme, the probability reaches a value of 27 % for NC and 25% for HC for a number of users equal to 120 compared to a blocking probability of 48% and 45% for NC and HC, respectively with the scheme CAC proposed in [10]. In Figure 4, we can observe that the application of our CAC scheme improves the values of blocking probabilities for two types of calls (HC and NC) for the NGBR traffic. In fact, using the proposed CAC scheme, the blocking probabilities reach the order of 32% and 30% for NC and HC respectively. while in [10], the achieved rates are 51% and 47% for NC and HC respectively.

Comparing between results in Figures 3 and 4, it is clear that the HCDP and NCBP values for the GBR traffic are lower than the HCDP and NCBP for the NGBR traffic. This is expected and can be explained by the introduction of the priority notion between the various service classes in terms of latency tolerance ($D_{max_{GBR}}$and $D_{max_{NGBR}}$). Moreover, the NCBP of the GBR traffic reaches higher values than the HCDP and this is can explained by the priority given to HC over NC in admission decision.

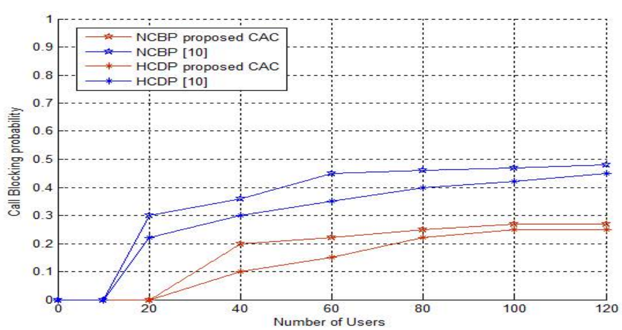


Figure.3.New call blocking/Handoff call dropping probability for GBR Traffic

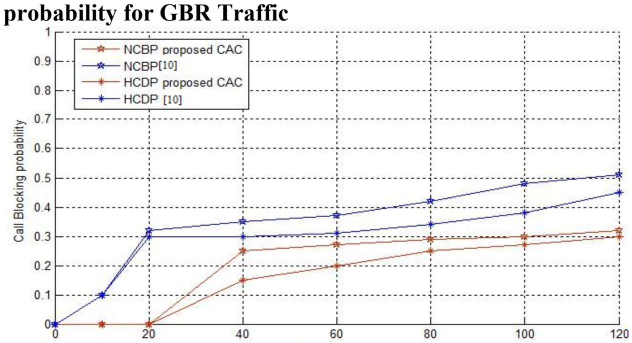


Figure.4.New call blocking/Handoff call dropping probability for NGBR Traffic

IV.2.2 Physical resource blocks utilization

The physical resource blocks utilization is the ratio of the number of allocated PRBs for the users in the system during the whole simulation time. The result of the PRB utilization according to the number of the UEs is shown in Figure 5.

If we apply our CAC scheme, the PRB utilization can achieve 96% whereas this value is only 75% in the CAC method defined in [10]. This gain (of about 21%) is observed for simulations involving more than 120 UEs. The best use of the PRBs is due to the concept of resource allocation algorithm, which adjusts the allocation of resource intelligently.

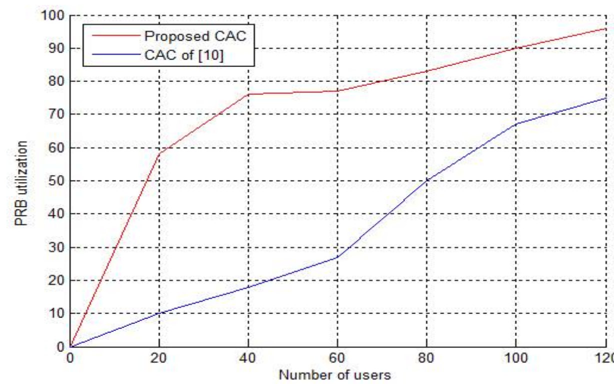


Figure.5.PRB utilization

IV.2.3 Served users

The results of the served number versus the total number of users as shown in Figure 6. It is clearly observed that the RUPSA scheme serves an interesting number of users. This is because the RUPSA adjusts the allocation of resource adaptively. Indeed, the RUPSA can schedule more users by giving the needful PRBs for each one. This allows accepting much more number of users and maximizing the total number of used PRBs.

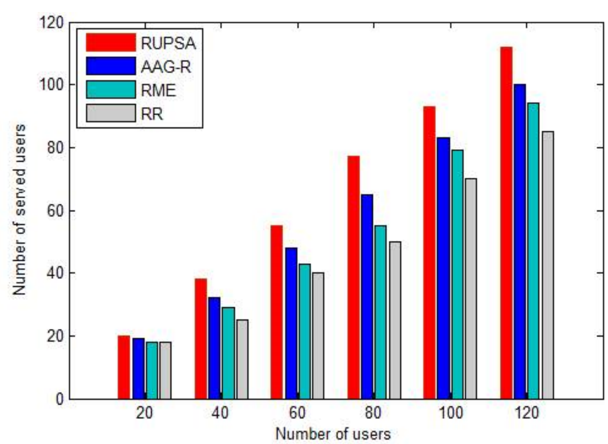


Figure.6.Served users

IV.2.4 System Throughput

Figure 7 shows the average system throughput of RR, RME, AAG-R and RUPSA algorithms as a function of the number of users. As we explained in the previous subsection, the RUPSA scheme can serve much more number of users compared to others algorithms. Serving more users requires harness the maximum of available resources blocks which increasing the overall throughput. In addition, RUPSA use the equation (7) to distinguish between users (less or more prioritize). On comparing with AAG-R scheduler, it is observed that RUPSA scheduler achieves highest throughput. For The RME, PRBs are more likely to be assigned to users with higher CQI values. But, most of the PRBs are wasted. On the contrary, UEs with lower CQI values can only transmit at low data rate because they get only very few PRBs. The RR algorithm is in fourth position. This is expected because neither the user requirement nor the channel quality is considered by the RR.

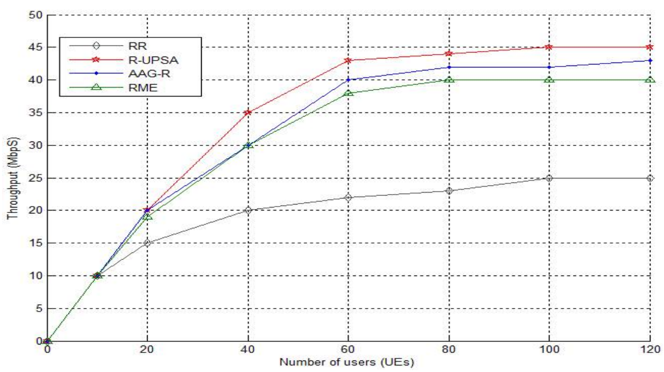


Figure.7. System throughput

IV.2.6 Fairness index

The fairness of the approaches was evaluated by the Jain's fairness index. The definition of this index is stated in [28],[25]. We can also calculate this fairness index as:

$$\boldsymbol{F}(\mathbb{C}_1, \mathbb{C}_2, \dots \mathbb{C}_n) = \frac{(\sum_{j=1}^{n} \mathbb{C}_j)^2}{\boldsymbol{n} \times \sum_{j=1}^{n} (\mathbb{C}_j)^2} \qquad \textbf{(9)}$$

Where n represents the total number of UEs and $\mathbb{C}_j$ is the number of resources assigned to user j. Jain's fairness index returns a value between 0 and 1. Value 1 represents the best fairness in the system.

Figure 8 shows the fairness results for the schedulers RR, AAG-R, RME and RUPSA. The maximum of Jain's fairness index is obtained by the RR scheduler. This is logical because RR assigns almost the same number PRBs for all UEs. Moreover, we observe that RUPSA achieves interesting results. This is explained by the fact that RUPSA serves the users according to their priorities as shown in equation (7). This equation contains factor $\mathrm{F_i}$ that provides the fairness among users. AAG-R and RME are in third and fourth position, respectively. The users that will receive resources are those with the best channel conditions.

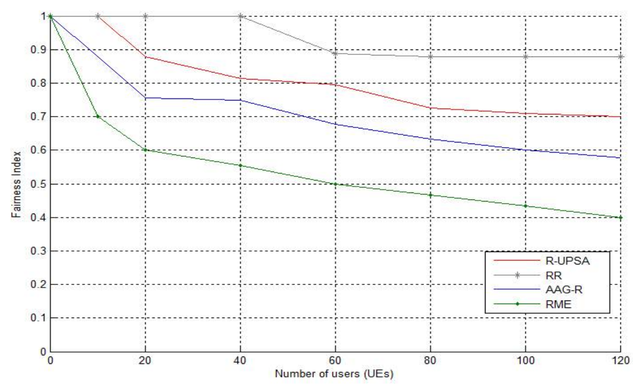


Figure.8.Fairness index

V. CONCLUSION

In this paper, we proposed some new CAC and scheduling algorithms. The CAC scheme aims to handle the NC and HC and the scheduling scheme aims to maximize the systems throughput, assign a fair distribution of PRBs and handle GBR and NGBR traffic in LTE Uplink systems. The performance of these algorithms was evaluated, considering LTE configuration parameters. The Simulation results show that the proposed schemes perform well in terms of the obtained a low dropping and blocking probability, system throughput, fairness index, served users and delay.

REFERENCES

[1] 3GPP, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (3GPP TS 36.213 version 8.8.0 Release 8)," pp. 1–79, 2009.

[2] 3Gpp, "Policy and Charging Control Architecture (TS 23.203 Release 11)," p. 38, 2012.

[3] H. Ekström, "QoS Control in The 3GPP Evolved Packet System," *IEEE Commun. Magazine.*, vol. 47, no. 2, pp. 76–83, 2009.

[4] M. Anas, C. Rosa, F. D. Calabrese, P. H. Michaelsen, K. I. Pedersen, and P. E. Mogensen, "QoS-aware single cell admission control for UTRAN LTE uplink," *Vehicular Technology Conference(IEEE VTC spring)*, vol. 1, no. 1, pp. 2487–2491, 2008

[5] 3GPP, "Way Forward on Power Control of PUSCH," R1-073224 , 2007

[6] H. Lei, M. Yu, A. Zhao, Y. Chang, and D. Yang, "Adaptive connection admission control algorithm for LTE systems," *Vehicular Technology Conference(IEEE VTC spring)* , pp. 2336–2340, 2008

[7] F. De Rango, A. Malfitano, and S. Marano, "GCAD: A Novel Call Admission Control Algorithm in IEEE 802.16 based Wireless Mesh Networks," *Journal of Networks*,

vol. 6, no. 4, pp. 595–606, 2011

[8] S. Ben Rejeb, “Admission Control Strategies and QoS Evaluation Based on Mobility in LTE,” *GLOBECOM,* pp. 4720–4725, 2013.

[9] J. Wang and Y. Qiu, “A New Call Admission Control Strategy for LTE Femtocell Networks,”*International Conference on Advances in Computer Science and Engineering,* pp. 334–338, 2013.

[10] F. Zarai, K. Ben Ali, M. S. Obaidat, and L. Kamoun, “Adaptive call admission control in 3GPP LTE networks,” *International Journal of Communication Systems Wiley*., pp. 1522–1534, october 2014.

[11] O. A. Ojesanmi and R. F. Famutimi, “Adaptive Threshold Based Channel Allocation Scheme for Multimedia Network,”*International Journal of Computer Science and Network Securit*y, vol. 9, no. 1, pp. 260–265, 2009.

[12] T.O Oyebisi, O.A Ojesanmi, “Development of Congestion Control Scheme for wireless mobile networks,” *Journal of Theoretical and Applied Information Technology*pp., 965–972, 2008.

[13] N. Omheni, A. Gharsallah, F. Zarai, M. S. Obaidat, and F. “Management Approach for HWNs,”*GLOBECOM*, pp. 112–117, 2014.

[14] L. Ruiz de Temino, G. Berardinelli, S. Frattasi, and P. Mogensen, “Channel-aware scheduling algorithms for SC FDMA in LTE uplink,” *International. Symposium. Personal, Indoor and Mobile Radio Communications(PIMRC)*, pp. 1-6, 2008

[15] F. Liu, X. She, L. Chen, and H. Otsuka, “Improved Recursive Maximum Expansion Scheduling Algorithms for Uplink Single Carrier FDMA System,” *Vehicular Technology Conference (VTC Spring)*, pp. 1–5, 2010.

[16] S. a. Alqahtani and M. Alhassany, “Performance Modeling and Evaluation of Novel Scheduling Algorithm for LTE Networks,” *Int. Network Computing and Applications (NCA).,* pp. 101–105, 2013.

[17] R. Ruby and V. Leung, “Towards QoS assurance with revenue maximization of LTE uplink scheduling,” *Communication Networks and Services Research Conference (CNSR), pp. 202–209, 2011.*

[18] O. Delgado and B. Jaumard, “Scheduling and resource allocation for multiclass services in LTE uplink systems,” *Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 355–360, 2010.

[19] M. Al-Rawi, R. Jäntti, J. Torsner, and M. Sågfors, “Opportunistic uplilnk scheduling for 3G LTE systems,” *Innovations in Information Technology,* pp. 705–709, 2008.

[20] R. Khdhir, K. Mnif, and L. Kamoun, “New scheduling algorithm for Uplink LTE System,” *International Symposium on Networks, Computers and Communications,* pp. 4–8, 2014.

[21] N. Prasad, H. Zhang, H. Zhu, and S. Rangarajan, “Multi-user MIMO scheduling in the fourth generation cellular uplink,” *Wireless Communications, IEEE Transactions on.*, pp. 1855–1859, 2013.

[22] S. Chaudhuri, I. Baig, and D. Das, “Utility based QoS aware Uplink Scheduler Scheme for LTE Small Cell Network,”*IEEE International Conference on Communications (ICC),* pp. 3149–3154, 2015.

[23] A. Kanagasabai and A. Nayak, “Opportunistic Dual Metric Scheduling Algorithm for LTE Uplink,” *International Conference on* Communication *Workshop(ICCW)*, pp. 1446–1451, 2015.

[24] F. C. Kuo, H. C. Wang, K. C. Ting, C. C. Tseng, and P. E. Liu, “Robust LTE uplink scheduling based on call admission control,” *International Conference on Convergence (ICTC)* , pp. 548–552, 2012.

[25] N. Abu-Ali, A. E. M. Taha, M. Salah, and H. Hassanein, “Uplink Scheduling in LTE and LTE-Advanced: Tutorial, Survey and Evaluation Framework,” *IEEE Commun. Survey and Tutorials*, vol. 16, no. 3, pp. 1–27, 2013.

[26] R. Kwan, R. Arnott, and M. Kubota, “On Radio Admission Control for LTE Systems,” *Vehicular Technology Conference Fall (VTC)*, no. 1, pp. 1–5, 2010.

[27] 3GPP, “Technical Specification Group Services and System Aspects; Technical Specification Group working methodsTR 21.900 *,Release 8,* pp. 1–6, *2010*.

[28] R. Jain, D.-M. Chiu, and W. R. Hawe, “A quantitative measure of fairness and discrimination for resource allocation in shared computer system,”*ACM Transaction on compter systems*. pp. 1–38, 1984.

A Facebook Identical Data Detection and Deletion Algorithm

Harshita Shukla
Dept. of Computer Engineering and Applications
National Institute of Technical Teachers' Training and Research (NITTTR), Bhopal (M.P), India.

Shailendra Singh, Senior Member IEEE
Dept. of Computer Engineering and Application
National Institute of Technical Teachers' Training and Research (NITTTR), Bhopal (M.P), India.

***Abstract*— Facebook is becoming very popular as millions of users are sharing their thoughts by using various data formats. The motive behind its launch was to find old friends and relatives and make new friends. All Social Networks need to meet the increasing user demands of data storage and retrieval. The Social Networks are based on cloud to deal with dynamic speed of data generation. The success of Facebook has resulted in increased user traffic and large amount of data is continuously generated by its users'. It requires novel ways of storing data and removal and removal of duplicates as much as possible while maintaining the speed of responding to a query. In this paper, an attempt is made for the identification of data duplication and its removal. Social networking sites need dynamic data management by identifying duplicate data and its deletion technique. The removal of duplicate data is necessary, not only to reduce runtime, but also to improve search accuracy and efficiency. The implementation of this method reduces the indexing time to a great extent by decreasing the collection length, resulting in the reduction of the amount of hardware required to support the system.**

Keywords- Hashing; indexing; similarity checking; unique documents; detecting replicate; data duplicity; web mining; Facebook.

I. INTRODUCTION *(HEADING 1)*

In 1990 the development of information exchange over the internet is rapidly increasing after the evolution of World Wide Web. Latest technological improvements in World Wide Web have empowered social interactivity through online communication. This communication and interaction among people who stay in geographically distinct locations has been possible through the evolution of online social networks. An online social network is a Web based communication service made available by various service providers to its users. social networks allows its user to make friends with known and also unknown people, share thoughts, pictures, images, other activities and information they like, play games, like each other's information's shared on their network etc. show in Fig. 1. A user can avail these services only after registering in a particular social networking site. In the process of registration, a user makes a virtual profile of himself in that website's domain using their Email Id's. The user profiles consist of their image, information related to their personal attributes as well as their interested areas and likings. Facebook, Twitter, LinkedIn are few of the big players in the area of online social networks.

The evolution of online communication using the internet has become one of the most popular areas of research. Many researchers have written about online social networking in their own ways. One of the definition of a social networking website is, it is a collection of millions of user profile connected with each other due to a relationship that may be friends, collogues, family members, community members etc. The topological view of these websites depicts a Graph like structure in which the user profiles can be considered as nodes and the relationships among various user profiles can be treated as links between them. So it can be said that a social networking website depicts a social graph comprised of user profiles and generic levels of interdependencies among them show in Fig. 2.

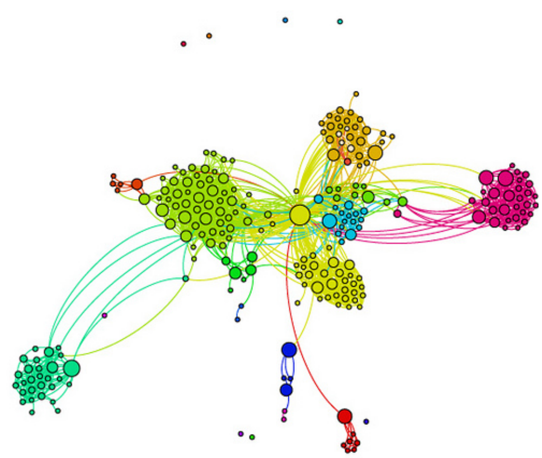


Figure 1. Data Sharing in Social Network.

Duplicity removal is a new technology with great significance in the area of online social networks because its real time application in any online social network can help them in reduction of their datacenter requirements. This objective can be achieved by reducing the capacity needed to store same amount of data on the disk in lesser space. Data duplicity removal works in two ways:

1. Content aware
2. Block level

Our algorithm uses the content aware technique because we use Facebook files as input dataset and the comparison is performed between a file pair. Unique codes are generated for each and every file and stored in an index table. The unique codes of these files are compared for identical file selection and then removal from the Facebook database servers.

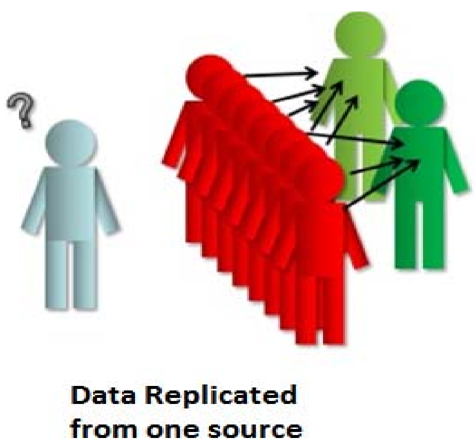


Figure 2. Data Sharing in Social Network.

The technique of identical data removal takes a dataset as input, make smaller chunks out of this set of input data. The size of chunks are still in kilobytes so these smaller chunks will be processed to examine at the file level to keep only unique data files on the disk. The unique files are selected by generation of unique codes based on hashing techniques.

One more benefit of this technique could be the lower network bandwidth need for data transfer because the basic concept of social networking websites are to upload and share data from multiple sources. The popularity of social networking websites depends on its user satisfaction, so higher network data transfer capacity becomes essential while Sharing data from multiple sources and remote location because it can significantly affect the popularity of any social networking websites and can cause severe harm to its business.

Facebook [9] is the largest social networking website to date with more than 950 million user profiles worldwide ingesting 500 terabytes of data into the database every day. The novel solutions provided by these websites for information exchange such as uploading a photo or video, clicking a notification, checking out a friend's link or visiting a page also results in duplicate content generation because, as per a statistics 2.5 billion Facebook's data items are shared per day by its users. The management of these data clusters requires thousands of storage devices and an efficient and dynamic indexing technique to deal with this giant ever changing Facebook database. Hence the company needs to figure out an efficient technique for analyzing its data and constantly keep a track on removal of duplicity in its database. In this research paper, we have proposed a technique to achieve above stated objective forn the Facebook web database.

II. Related Work

The social networking websites [1] allow its users to share data with each other in their friends list and also to other connected user through a community relationship. Many users staying at different geographic locations can share the same information like any news about a current issue, quotations, images, videos etc. in their profile pages and also on any community page they are joined to.

The users are fond of reposting the same content several times in their timeline. This activity generates a large amount of duplicate data in social network database servers. It has also come in existence that identical documents are becoming silent but very serious threat for community database servers as it affects the search and execution time of query. The classification of duplicate content is important because in textual data some natural text duplicates can occur as a result of a language's phenomenon.

In as it is reposted data will be considered as identical and if reposted with slight changes then considered as near identical. These co-derivative contents will be taken care of in this dissertation work. The Face book database is a huge one and also dynamic in nature so identification and deletion of duplicate content from it, is not an easy task. Perhaps it will be prohibitively expensive to compare each file pair of every individual node of the social graph. Sequential compression, delta encoding and jacquard's similarity check are some of the identical detection methods which are used to eliminate redundancy in datasets. Content based identical document detection can be carried out at one of the three levels of granularity:

1. Whole file
2. Fixed size blocks
3. Variable size chunks

These file size distributions are generated by a content defined chunking algorithm. "Fingerprint" method is also used in detecting identical documents. "Fingerprints" can also be called as document checksum which is an output of a hash function such as MD5. Use of checksum is advantageous because it reduces the actual size of the original database resulting in cheaper cost of document comparison while preparing the Index [2].

As per the index hashing technique two duplicate documents will have the same hash value whereas the non-identical documents will have a very high probability of having a different checksum value. Moreover it's easy to store hash values in memory [3]. Detection of identical document can then be performed easily and then these duplicate contents can be removed with a single pass over the dataset by comparing the hash values of both the documents against each other.

Identification and deletion of near duplicate documents is more difficult because a small change (e.g. a single byte) in the input value can affect the output completely [4]. So for near identical documents simple hashing algorithm will not work as efficiently as it works for completely identical documents. Hence the web application that needs to find out near identical documents requires special hashing techniques.

Some of the most popular techniques are overviewed in this section. Broder's shingling algorithm [5] represents each document as a set of shingles (K-grams). These shingles are a sequence of any k consequent words. For example: Let, there is a document D then, SD is a set of all shingles that occur in D Now the document similarity measure is computed as-

$R(x,y)=|Sx \cap Sy|/|SxUSy|$

This is Jacquard's similarity measuring formula. Here the resemblance value results in the interval [0, 1]. The document with higher similarity measure will be close to 1 and the document with lower similarity level will be near to 0. Broder's algorithm says that exact resemblance value is not required to decide the similarity of a document pair. A predefined threshold value is used for comparing with the document resemblance value. Any value above this threshold value, suffice the document resemblance. The identical files can easily be approximated accurately with the use of small sample of the shingles which results a significant amount of reduction in the computational cost, because there are fair chances of selection of same shingle from a near identical document pair if the shingle selection is done on the basis of a specific feature such as lowest hash value.

Chowdhury et al. [6] utilized multiple data collections to evaluate the performance of their proposed algorithm called I-Match. The employed document collections vary in document lengths, size, and degree of expected document duplication. They used NIST and Excite at the Home as the data source. The I-Match algorithm illustrates that the input data operates on the basis of number of documents and it deals with documents of all sizes efficiently. Their method proved to have improved accuracy of duplicate detection in comparison with the state of the art methods and the execution time was about one-fifth of the time taken by other algorithms.

Šimon Suchomel [7] describes an architecture and concepts of a real-world document retrieval system, which is a part of a general anti-plagiarism software. Up to date systems for plagiarism detection are discussed from the source retrieval perspective. The key approaches of source retrieval are compared. The system recommendations stem from design, implementation, and several years of operation experience of a nationwide plagiarism solution at Masaryk University in the Czech Republic. The design can be adapted to many situations. Proper usage of such systems contributes to the gradual improvement of the quality of student theses.

In high tech white paper [8], Ravindra Mahabaleshwar has described that IT companies are keen to develop a methodology to reduce the maintenance cost for infrastructure management. With the exponentially growing data sizes in enterprise domain, this can be achieved by reducing the number of data centers required to cater their database needs. The challenging part of this task is to maximize the data compression ratio with negligibly affecting throughput. Data de-duplication technique is helpful in achieving this objective of storing same data volume in lesser storage space but as the operation is resource sensitive it can badly affect the enterprise business, if implemented incorrectly.

To achieve de-duplication, all the database files are compared with each other by following various steps and if any identical content found in a pair of files then it is removed from the database. This can provide 2 to 200 times space storage space reduction and low bandwidth data transfer. The process of identical removal has four basic steps- data segmentation, index generation, comparing the index values of two files for identical detection and storing only unique files on storage disk. In this paper, various techniques of implementing the de-duplication method, their effect on data de-duplication ratio and throughput is discussed in detail.

III. PROBLEM STATEMENT

Web based databases are critical in handling if content duplicity occurs in it because it may be possible that all the results returned by a given query shows identical documents. This phenomenon decreases the usability of the database and also reduces the speed. In case of Facebook social networking site a huge amount of duplicate data generated because its users tend to repost the same content in their timeline and also in other's page several time weather it may be a quotation, image, and video. Most of the memory space in the Facebook database server is occupied with duplicate data. The query processing time and searching time is greatly increased without any fruitful result and hence reducing the productivity of the web application. To achieve high level user satisfaction, it is important to detect and remove identical and near identical data from the social network. As we know, that Facebook is the most popular social media now-a-days, so content redundancy checking and removal becomes the top most priority for it to achieve high level user satisfaction.

IV. PROPOSED WORK

For duplicate document detection, this technique uses an indexing method based on secure hash algorithm with a differently favorable strategy for Facebook database. It is desirable to design an algorithm that is not required to solve any hard sub problem but can give nearly optimal solutions for data clustering. This method can obtain optimal solutions quicker via differently favorable strategy.

In this technique first of all, scrutinize the files of a Facebook node into several categories such as small files, html files, audio file, video file, image files etc. so that the process of comparison becomes well organized and faster.

The algorithm will first sort data entries in lexicographic order which brings all the identical entries next to each other in a sequence. In this process each comparison will either delete a record if found duplicate or move to next entry in the database.

Then, after the completion of each comparison, a sorted list of unique records is generated.

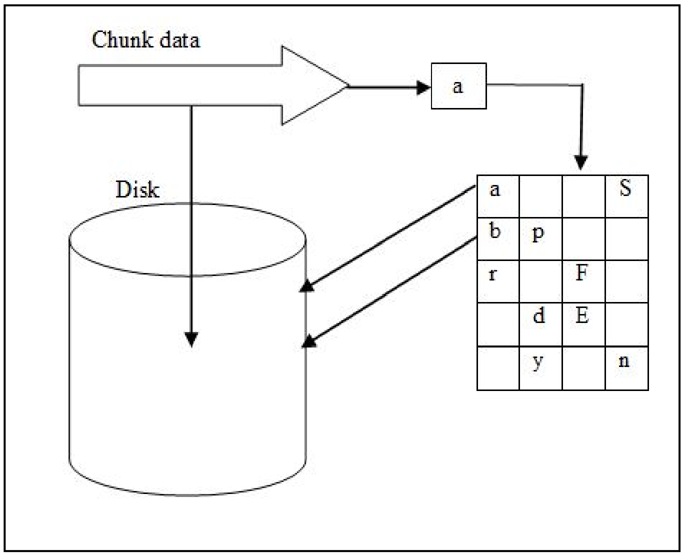


Figure 3. Creating Unique Identifier and index

The secure hash algorithm is of 4 types: SHA-1, SHA-256, SHA-384, and SHA-512. All these algorithms uses almost similar function with a few different descriptions and they take different sized words as input. The different word size for each algorithm is mentioned in Table I.

TABLE I. MESSAGE DIGEST SIZE OF 4 TYPES OF SECURE HASH ALGORITHM

Hash Algorithm	Size of digest
MD5	128
SHA1	160
SHA256	256
SHA512	512

To produce identical results, each of these algorithms includes:

a. *Ch(a, b, c)* and *Maj(a,b, c)* functions;

b. Bitwise OR *operation (U).*

The MD5 message digest algorithm is widely used in cryptographic hash function to check data integrity and it produces a 128 bit hash value. It is expressed in hexadecimal number of 32 digits. This algorithm takes an arbitrary length message as input to produce a 128-bit hash value as output that is to be stored in an Index table. These hash values are unique because two different input messages cannot produce same hash value due to its computational infeasibility. All these secure hash algorithms are divided in 2 stages

1. Preprocessing: It includes padding a message, parsing the padded message into m-bit blocks and setting the initial hash value.
2. Hash computation: this operation is performed to iteratively generate a series of hash values.

The messages are padded to ensure the size of the message which should be in multiple of 512 or 1024 bits according to the algorithm, then in the next step these messages are parsed into N; m-bit blocks before the beginning of hash computations.

The SHA-1 uses a sequence of logical function and SHA-256 uses six logical functions. They both operate on 32-bit words and produce a new 32-bit word as output. Whereas the SHA-384 and SHA-512 uses 6-logical function and each function operates on 64-bit words and results in a new 64-bit word.

The SHA-256 function quickly compares large number of files and generates exclusive hash values for each entity entries since the chance of two dissimilar files having the identical hash value is extremely isolated.

This de-duplications algorithm is a technique of removing duplicate content from Facebook database servers. This technique is based on indexing approach of secure hash algorithm in which unique identifiers are generated by converting the data objects of the available file into binary codes. These binary codes will always generate unique values for each file show in Figure. 3.

Method is implemented as described in the following steps:

Algorithm

Input: Feature set FDbi = Facebook Database FDbi Through Searching Cloud Database,

---

Output: $Rp_{n-1}$for Deleting and FDb_i as a Unique Data.

begin

1. Upload Facebook Database FDb_i Through Searching Cloud Database.
2. For each Record of Facebook Database FDb_i, Create index Pointer SHA_i which points to FDb_i .
3. Let *Fbx* and *Fby* be two Facebook Contents.
4. Now two files *Fbx* and *Fby* will be selected for comparison from Facebook database.
5. Arrange the index values of file *Fbx* in lexicographic order.
6. The data object of file *Fbx* will be compared from *Fby* through given equation

$$U_d(Fb_x, Fb_y) = \frac{\sum_{t_i \in S(Fb_x) \cap S(Fb_y)} w(t_i)}{\sum_{t_j \in S(Fb_x) \cup S(Fb_y)} w(t_j)}$$

Where $w(t_i)$ is the weight of t_i assigned by the hash function scheme.

7. As for a predefined threshold θ, if Ud(Fbx,Fby)> θ then text *Fbx* and *Fby* are considered as duplicates.
8. Repeat Step 3 to Step 8 until $FDb_i = FDb_{n-1}$

where i= {1,2, ……n-1}

9. Return Rp_{n-1} for Deleting and FDb_i as a Unique Data.

End

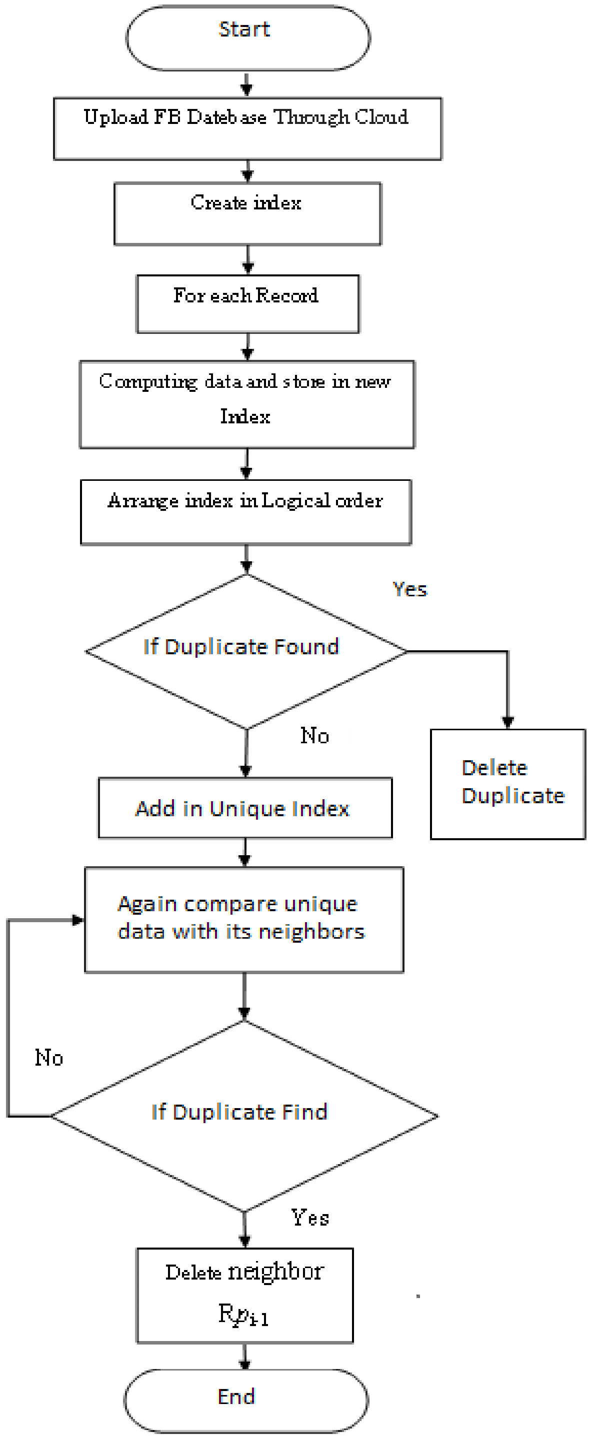


Figure 4. Flowchart for Creating Unique Identifier and Index

For experimentation purposes 5 different Facebook accounts are taken and connected with each other in friend list shown in table 2. The proposed technique uses cut-off thresholds to filter any word above and below certain normalized idf values. Document sizes both pre and post filtration and the timing results are collected. Document size information is used to determine the sensitivity of these types of algorithms to smaller documents.

This method is accomplished with the following methodology that is adapted in Fig. 4.

Duplicate content is identified by calculating similarity between the two data. Data with higher duplicate value in Comparison to a predefined threshold is considered as duplicates. Let there be two Facebook Contents, the standard duplicate value is defined as follows:

$$U_d(Fb_x, Fb_y) = \frac{|S(Fb_x) \cap S(Fb_y)|}{|S(Fb_x) \cup S(Fb_y)|} \quad (1)$$

where |S| denotes the size of set S. In equation (1), all terms are considered of equal importance. This may be unsuitable because different terms have different meaning. The weighted duplicate value can be defined as follows:

$$U_d(Fb_x, Fb_y) = \frac{\sum_{t_i \in S(Fb_x) \cap S(Fb_y)} w(t_i)}{\sum_{t_j \in S(Fb_x) \cup S(Fb_y)} w(t_j)} \quad (2)$$

Where $w(t_i)$ is the weight of t_i assigned by the hash function scheme. As for a pre defined threshold θ, if $U_d(Fb_x, Fb_y) > \theta$, then text Fbx and Fby are considered as duplicates.

Here using general hash function (e.g. SHA-1, MD5) which are designed to make the hash values be uniform distribution as possible. Perceptibly, a small disparity between different documents will get a fairly different hash value. As a matter of course, we need a hash function that will get similar values when the inputs are similar.

In this technique first of all we scrutinize the files of a Facebook node into several categories such as small files, html files, audio file, video file, image files etc. so that the process of comparison will become well organized and faster. Table 3 gives a representation of different file types and number of documents found for each file type from a single node of the Facebook social graph.

Table 2 contains the notational description of the proposed Algorithm technique experiments. We used four document collections, as shown in Table 2. Each collection was chosen to test particular issues involved with identical detection. The first is a Harshita Shukla Facebook document collection flagged as Identical. Only 5 Facebook users can give their database due to privacy and security reason. But from this database we can easily calculate an idea of identity document.

The data collection for this research was produced from 5 user documents. This data were then filtered by the Facebook

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

account holders to include only those documents thought to be 'Identical.' The collection contains 977 documents, each of which is suspected of having an identical web document within the collection. Many titles are in the collection repeatedly because of multiple spider inputs. This collection is approximately 165 megabytes in size. The Facebook collection is highly identical. Thus, as better approaches are used, the greater is the percentage of the collection found as identical.

TABLE II. EXPERIMENTAL COLLECTIONS

S.No	Facebook Username	Size of documents (MB)	Number of collection
1.	Harshita Shukla	13.5	161
2.	Niranjan Singh	5.5	102
3.	Ankita Shukla	39.4	31
4.	Pradeep Singh	10.5	191
5.	Ravi Singh Baghel	95.2	483
	Extra	1.11	9
	Total	165.21	977

TABLE III. CLASSIFICATION OF FACEBOOK'S FILES TYPES

File Type	Total number of documents
Small file	15
Html file	10
Picture file	1423
Audio file	10
Video file	132
Other file	46

Many titles are in the collection repeatedly because of multiple spider inputs. This collection is approximately 165 megabytes in size. The Facebook collection is highly identical. Thus, as better approaches are used, the greater is the percentage of the collection found as identical.

The effect of filtering tokens on the degree of identical document detection is shown in Table IV. The percentage of Identical found is an evaluation metric of the effectiveness of the filter. Also shown in the table, is the percentage of terms retained after each filtering technique. As shown in Table 4, the higher the filtration, the greater the degree of detection.

From the Fig. 5 and Table 4 it is clear that in Facebook there is 20% image and 25 % video file are identical. Our simple filtering techniques reduced the list of tokens used to create the hash. By eliminating white spaces and only keeping unique tokens, many small document changes are eliminated. Keeping only unique tokens eliminates movement of paragraph errors, stemming removes errors caused by small token changes, and stop word removal removes errors caused by adding or removing common irrelevant tokens, in terms of semantics.

This algorithm will first perform sorting of data entries in lexicographic order this will bring all the identical entries next to each other in a sequence. In this process each comparison will either delete a record as a duplicate found or move to next entry in the database. Then, after the completion of each comparison, a sorted list of unique records will be generated.

TABLE IV. IDENTICAL DOCUMENTS AND PERCENT FOUND AS IDENTICAL FOR SMALL, HTML, IMAGE, AUDIO, VIDEO AND OTHER FILE

File type	Found as identical (%)	Identical documents found in collection	Total no of documents
Small file	0%	0	1
Html file	0%	0	15
Picture file	19.18%	177	923
Audio file	0%	0	0
Video file	25%	8	32
Other file	0%	0	6

TABLE V. SHOW COMPUTING RESULT WITH TIME ELAPSE AND FILE SIZE

Total files found	Sum of file sizes (MB)	Elapsed time for search (sec)	Computing hashed files	Sum of file sizes now (MB)	Elapsed time for computing hashes (sec)	Deleted files	Elapsed time for deleted (sec)
977	165.21	2.91	186	84.18	2.21	185	2.56

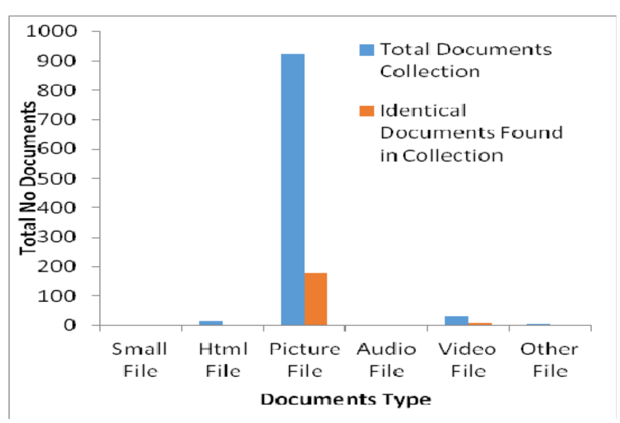


Figure 5. Identical Documents and Percent Found

v. conclusion

At present, Facebook is trying to manage its growing armada of servers and also looking for new ways to improve the scalability of its data center infrastructure. This research is aiming to design and develop an efficient method for identical data detection which can easily identify identical in Face book and online social network database.

This technique cleans up all the duplicate content from database center in cloud server by managing the entire Face book database for efficient storage utilization. It will help Facebook data centers to operate on as little ongoing maintenance as per possible in future, even if data growth rate is exponential. This approach detects similar data which is of critical importance in applications where data is obtained from social media. The removal of similar data is necessary, not only to reduce runtime, but also to improve search accuracy. Reduction of the collection sizes results in great savings in indexing time and a reduction of hardware requirement to support the system.

References

[1] F. Schneider, A. Feldmann, B. Krishnamurthy, W. Willinger, “Understanding online social network usage from a network perspective”, In Proceedings of the ACM SIGCOMM Conference on Internet, Measurement, PP 35–48, 2009.

[2] Henzinger, M., “Finding near-duplicate Web pages: a large-scale evaluation of algorithms.” SIGIR, PP 284-291, 2006.

[3] Yi Yu, Zijian Huand Yuzhu Zhang, “Rearch on Large Scale Documents Deduplication Technique based on Simhash Algorithm”. International Conference on Information Sciences, Machinery, Materials and Energy ICISMME, 2015.

[4] Charikar, M.S., “Similarity estimation techniques from rounding algorithms”, STOC, PP 380-388, 2002.

[5] A.Z. Broder. “Identifying and filtering near-duplicate documents”. Proceedings of the 11th Annual Symposium on Combinatorial Pattern Matching, pages 1–10, 2000.

[6] Chowdhury, A., at.el, “Collection statistics for fast duplicate document detection.” ACM Trans. Inf. Syst., Vol 20(2), PP 171-191, 2002.

[7] Šimon Suchomel and Michal Brandejs, “Source Retrieval for Plagiarism Detection”, Journal of Advances in Information Technology Vol. 6, No. 1, February 2015.

[8] HiTech Whitepaper Effective Data Deduplication Implementation 05 2011.

[9] https://gigaom.com/2012/08/22/facebook-is-collecting-your-data-500-terabytes-a-day/

[10] Shivakumar, N., “SCAM: A copy detection mechanism for digital documents” In DL, 1995.

[11] Rivest, R., "The MD4 Message Digest Algorithm", RFC 1320, MIT and RSA DataSecurity, Inc., April 1992.

[12] “The SHA-1 algorithm specified in this document is identical to the SHA-1 algorithm”, FIPS 180-1.

[13] Shivakumar, N. and Garcia Molina, H., "Finding near-replicas of documents on the web", Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Vol. 1590, PP 204-212. 1998.

[14] T. Parimalam, R. Deepa, R. Nirmala Devi, P. Yamuna Devi, " Detecting Duplicate Records - A Case Study ", International Journal of Scientific Research in Science, Engineering and Technology, PP 290-294, 2015.

[15] Sarah Weissman, Samet Ayhan, Joshua Bradley, and Jimmy Lin ,“Identifying Duplicate and Contradictory Information in Wikipedia”.Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries ,Pages 57-60, 2015.

[16] Thorsten Papenbrock, Arvid Heise, and Felix Naumann,”Progressive Duplicate Detection”, IEEE TRANSACTIONS on knowledge and data engineering, vol. 27, no. 5, PP 1316 – 1329, 2015.

AUTHORS PROFILE

Authors Profile …

Rule Generation for Proton Pump Inhibitor Regimen Using Learning Vector Quantization and C4.5

Anifuddin Azis
Department of Computer Science and Electronics
Gadjah Mada University
Yogyakarta, Indonesia

Edi Winarko
Department of Computer Science and Electronics
Gadjah Mada University
Yogyakarta, Indonesia

Sri Hartati
Department of Computer Science and Electronics
Gadjah Mada University
Yogyakarta, Indonesia

Zullies Ikawati
Department of Pharmacy
Gadjah Mada University
Yogyakarta, Indonesia

Abstract—**The excessive or irrational use of drugs categorized as Proton Pump Inhibitor (PPI) was indicated in Baptis Hospital of Kediri, Indonesia. In the PPI-based drug regimen among patients with digestive disorders from December 2009 to February 2010, many cases that the PPI-based drug regimen was not in accordance with the prevailing procedures were found, i.e. the drug regimen among patients who should not be given it.**

In this study, a method was developed to generate the PPI-based drug regimen rule. Data on the PPI-based drug regimen were trained using Learning Vector Quantization (LVQ) algorithm. The results of LVQ were stored as new data, which were extracted into IF-THEN rule with C4.5 algorithm.

Based on the test, eighteen rules were generated for the PPI-based drug regimen with an accuracy rate of 82.5% on test data.

Keywords—PPI-based drug regimen; rule generation; LVQ; C4.5

I. INTRODUCTION *(HEADING 1)*

In Indonesia, the efforts of increasing the rationalization of drug use was determined in Decree of the Minister of Health RI Number 1197/MENKES/SK/X/2004 on the function and scope of hospital pharmacy services, i.e., among others, to review drug use in hospital by studying medical records compared with the standards of diagnosis and therapy. This review aims at continually enhancing the use of drug rationally.

The excessive or irrational use of drugs categorized as Proton Pump Inhibitor (PPI), consisting of omeprazole, lansoprazole, esomeprazole, rabeprazole, pantoprazole, was also indicated in Baptis Hospital of Kediri. In the PPI-based drug regimen for patients with a digestive disorder from December 2009 to February 2010, many cases were found, indicating that the PPI-based drug regimen was not in accordance with the prevailing procedures, i.e. the drug regimen for patients who should not be given it.

One of the ways to overcome such problem is to create a computerized system that stores a drug regimen rule. The rule is generally in a form of production rule (IF-THEN) with an advantage of being easily understood by users. In pharmacy, this method is largely applied [1][2]. However, the method has a common disadvantage if the knowledge required is not incomplete, inappropriate, and uncertain. In the method, no learning can be done [3].

On the other hand, there is an inductive learning system that makes a generalization from data/example data. Thus, the learning process do not require knowledge, but large data. Artificial neural network (ANN) is one of the empirical learning methods proved to be more excellent than or equal to other empirical learning ones in ability of making generalization [4]. ANN was successfully applied in various fields [5][6][7][8].

Learning Vector Quantization (LVQ) is one of the ANN methods to make a learning a supervised competitive learning. LVQ has a good accuracy rate and lower computation time than backpropagation [9].

In this study, a method was developed to generate the PPI-based drug regimen rule by extracting rules in data on the PPI-based drug regimen that were trained using LVQ algorithm. The extraction of rules was done with C4.5 algorithm. The method do not aims at enhancing the ability of LVQ. Instead, it employs LVQ as a pre-process for a specific rule induction approach, i.e. C4.5 Rule. Based on the test, the method successfully generated the PPI-based drug regimen rule with a good accuracy rate.

II. METHODOLOGY

A. Artificial Neural Network

Artificial Neural Network (ANN) is an information processing system that has a certain working characteristics identical with the human biological neural network working

system [10]. It was successfully applied in various fields, such as computer vision, image/signal processing, voice/character recognition, medical image analysis, remote sensing, industry inspection [11].

The advantages of ANN are [12]:

- ANN stores its knowledge in a form of weight, and using the weight value, it can make a simple and fast operation.
- It can operate with incomplete data and make a generalization well in similar data.

Despite an accuracy rate frequently better than that of other method, ANN is generally difficult to understand in generating relevant decisions with its complex architecture. Thus, it is often said that ANN is a "black box" method. Even, despite simple single architecture, it is still generally difficult to explain why a pattern include into a class or other patterns include in other classes.

To overcome such weakness, it is necessary to find a method in order that ANN can make explanation on the resultant conclusion, one of which is to extract ANN into IF-THEN rule [13].

B. Leraning Vector Quantization

Learning vector quantization (LVQ) [10] is a pattern classification method in which each output unit represents a particular class or category. (Several output units should be used for each class.) The weight vector for an output unit is often referred to as a reference (or codebook) vector for the class that the unit represents. During training, the output units are positioned (by adjusting their weights through supervised training) to approximate the decision surfaces of the theoretical Bayes classifier. It is assumed that a set of training patterns with known classifications is provided, along with an initial distribution of reference vectors (each of which represents a known classification). Figure 1 shows LVQ network with six units at the level of input and two units (neurons) at the level of output.

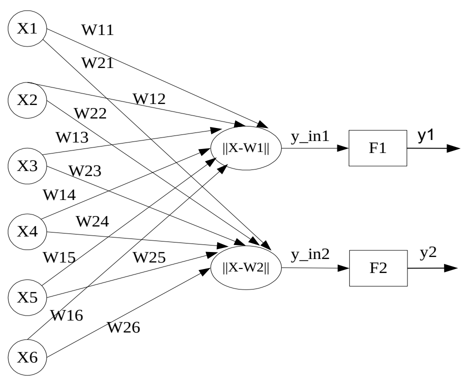


Figure 1. Example of the architecture of LVQ network.

The activation function of F1 will map y_in1 to y1= 1 if ||X-W1|| < ||X-W2||, and y2 = 0. Also, the activation function of F2 will map y_in2 to y2 = 1 if ||X-W2|| < ||X-W1||, and y1 = 0.

Kohonen [14] developed LVQ algorithm and named LVQ 2.1 that is a development of LVQ 2 and LVQ 1. The algorithmic step is to seek the nearest weight of each input data. If the category of input was the same with the category of wANNer or runner-up, the change in weight will be done.

x : current vector input.

y_{c1} : weight of wANNer/runner-up, category of y_{c1} is the same with category of x

y_{c2} : weight of wANNer/runner-up, category of y_{c2} is not the same with category of x

d_{c1} : distance between x and y_{c1}

d_{c2} : distance between x and y_{c2}

The calculation of "window" in LVQ 2.1 or improvement in weight of winner/runner-up (y_{c1} and y_{c2}) is done, if:

$$\min[(d_{c1} / d_{c2}), (d_{c2} / d_{c1})] > (1 - \varepsilon) \text{ and}$$

$$\max[(d_{c1} / d_{c2}), (d_{c2} / d_{c1})] < (1 + \varepsilon) \quad (1)$$

If the requirements were met, then:

$$y_{c1} \text{ (new)} = y_{c1} \text{ (old)} + \alpha (x - y_{c1} \text{ (old)}) \quad (2)$$

$$y_{c2} \text{ (new)} = y_{c2} \text{ (old)} - \alpha (x - y_{c2} \text{ (old)}) \quad (3)$$

If the requirements were not met, then improve the weight using LVQ 1. The change or improvement in code-book in LVQ1 as a basic process of LVQ is as follows:

if x and m_c are at the same class,

$$m_c(t + 1) = m_c(t) + \alpha(t)[x(t) - m_c(t)] \quad (4)$$

if x and m_c are at the same class,

$$m_c(t + 1) = m_c(t) - \alpha(t)[x(t) - m_c(t)] \quad (5)$$

if x and m_c are at the different classes,

$$m_c(t + 1) = m_c(t) \text{ for } i \neq c \quad (6)$$

m_c : the weight nearest to class x.

C. C4.5 Algorithm

Learning Generally C4.5 algorithm to build a decision tree is as follows [15]:

1) *Choose an attribute as root*
2) *Make a branch for each value*
3) *Share cases in the branches*
4) *Iterate the process for each branch until all cases in the branches have the same class.*

Choosing an attribute as root is based on the highest gain value of existing attributes. To calculate the gain, a formula is used as seen in (7).

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy(S_i) \tag{7}$$

{ $S_1, S_2, S_3, \ldots, S_n$} : partition S, in accordance with the value of attribute A

A : Attribute

n : Number of partitions of attribute A

$|S_i|$: Number of cases in partition S_i

$|S|$: Number of cases in S

Meanwhile, the calculation of entropy value can be seen in the following equation (4):

$$Entropy(S) = \sum_{i=1}^{n} -pi * \log \tag{8}$$

S : A set of cases

n : Number of cases in partition S

pi : Proportion of S_i to S

III. The Proposed Method

In the study, data on the PPI-based drug regimen were studied using LVQ that has the good ability of making generalization. Moreover, the results of LVQ learning will be induced to be a rule with C4.5, so that it is easier to understand as a guideline for the PPI-based drug regimen. Thus motivation behind the approach is to combine the generalization power of ANN and the easy understanding of rule. The method proposed is called as the formation of LVQ-C4.5 rule. The steps in the proposed method were presented in Figure 2.

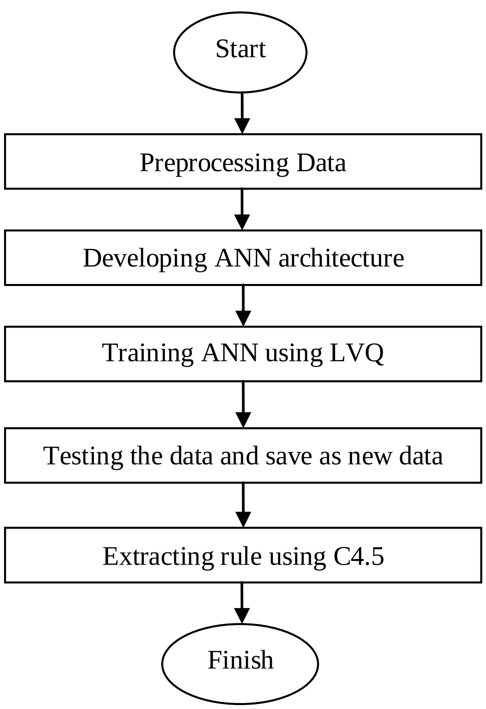


Figure 2. Flow chart of the steps of proposed method.

1) *Prepocessing Data*

Data of PPI drug regimen was taken from the data patient in questionnaire form. This data was processed and every symptoms from the patiens was listed and presented in Table I, if the symptom is present at that data it is defined by one (1) and if it is absent at that data it is defined by zero (0), presented in Table II. Regimen is class of data, if PPI drug regimen was correct to given to patient it is defined by one (1) and if PPI drug regimen was incorrect to given to patient it is defined by zero (0).

TABLE I. Symptoms from data patient

Code	Symptom name
A1	nausea/vomit
A2	stomachache
A3	abdominal bloating
A4	hot stomach
A5	heart pain
A6	Fullness
A7	swelling abdomen
A8	upper abdominal pain
A9	lower abdominal pain
A10	stomach cramp
A11	difficult to swallow
A12	Diarrhea
A13	abdominal pain up to waist

TABLE II. Prepocessing of PPI drug regimen result

Num	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	Regimen
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0	0	0	0	0
4	1	0	0	0	0	0	0	0	0	0	0	0	0	1
5	1	0	0	0	0	0	0	0	0	0	0	0	0	1
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	1	0	0	0	0	0	0	0	0	0	0	0
8	1	0	0	0	0	0	0	0	0	0	0	0	0	1
9	0	0	0	0	1	0	0	0	0	0	0	0	0	1
10	1	0	0	0	0	0	0	0	0	0	0	0	0	1
11	1	0	1	0	0	0	0	0	0	0	0	0	0	1
12	0	0	0	0	0	0	0	0	1	0	0	0	0	0
13	1	1	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	1	0	0	0	0	0	0	0	0	0	0
15	1	0	0	0	0	1	0	0	0	0	0	0	0	1
16	0	1	0	0	0	0	0	0	0	0	0	0	0	1
17	0	0	0	0	0	0	0	1	0	0	0	0	0	1
18	0	0	0	0	0	0	0	1	0	0	0	0	0	1
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	1	0	0	0	0	0	0	0	0	0	0
21	0	0	1	0	0	1	0	0	0	0	0	0	0	0
22	1	0	0	0	0	0	0	0	0	0	0	0	0	1
23	0	1	0	0	0	0	0	0	0	0	0	0	0	1
24	0	0	0	0	0	1	1	0	0	0	0	0	0	0
25	1	0	0	0	0	0	0	0	0	0	0	0	0	1
26	0	0	0	0	1	0	0	0	0	0	0	0	0	1
27	0	1	0	0	0	0	0	0	0	0	0	0	0	1
28	0	0	0	0	0	0	1	0	0	0	0	0	0	1
29	0	0	1	0	0	0	1	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	1	0	0	0	0	0	1
31	0	0	1	0	0	1	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	1	0	0	0	1
33	1	0	0	0	1	0	0	0	0	0	0	0	0	0
34	0	1	0	0	0	0	0	0	0	0	0	0	0	1
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

37	0	0	0	0	0	1	0	0	0	0	0	0	0	1
38	0	0	0	0	1	0	0	0	0	0	0	0	0	1
39	0	0	0	0	0	0	0	1	0	0	0	0	0	1
40	0	1	0	0	0	0	0	0	0	0	0	0	0	1
41	0	0	0	0	1	0	0	0	0	0	0	0	0	1
42	0	0	0	0	1	0	0	0	0	0	0	0	0	1
43	0	0	0	0	0	1	0	0	0	0	0	0	0	1
44	1	0	1	0	0	0	0	0	0	0	0	0	0	1
45	0	0	0	0	0	0	0	1	0	0	0	0	0	1
46	1	0	0	1	0	0	0	0	0	0	0	0	0	1
47	0	0	0	0	1	0	0	0	0	0	0	0	0	1
48	0	0	0	0	1	0	0	0	0	0	0	0	0	1
49	1	1	0	0	0	0	0	0	0	0	0	0	0	0
50	0	1	0	0	0	1	0	0	0	0	0	0	0	1
51	1	0	0	0	0	0	0	0	0	0	0	0	0	1
52	1	0	0	0	0	0	0	0	0	0	0	0	0	1
53	0	0	0	0	1	0	0	0	0	0	0	0	0	1
54	1	0	0	0	0	0	0	0	0	0	0	0	0	1
55	1	0	1	0	0	0	0	0	0	0	0	0	0	1
56	0	1	0	0	0	0	0	0	0	0	0	0	0	1
57	0	0	0	1	0	1	0	0	0	0	0	0	0	1
58	0	0	1	0	0	0	0	0	0	0	0	0	1	1
59	0	1	0	0	0	0	0	0	0	0	0	0	0	1
60	0	0	0	0	1	0	0	0	0	0	0	0	0	1
61	1	0	0	0	0	1	0	1	0	0	0	0	0	1
62	0	0	0	0	1	0	0	0	0	0	0	0	0	1
63	1	0	0	1	1	0	0	0	0	0	0	0	0	1
64	1	0	0	0	0	1	0	0	0	0	0	0	0	1
65	1	0	0	0	0	0	0	0	0	0	0	0	0	1
66	1	1	0	1	1	0	0	0	0	0	0	0	0	1
67	0	1	0	0	0	0	0	0	0	0	1	0	0	0
68	0	0	0	0	0	0	0	1	0	1	0	0	0	0
69	0	0	1	0	0	1	0	0	0	0	0	0	0	0
70	1	1	0	0	0	1	0	0	0	0	0	0	0	0
71	1	0	0	0	1	0	0	0	0	0	0	0	0	0
72	1	1	0	0	0	0	0	0	0	0	0	0	0	0
73	0	1	0	0	0	0	0	0	0	0	0	0	0	1
74	1	0	0	0	1	0	0	0	0	0	0	0	0	0
75	1	0	0	0	1	0	0	0	0	0	0	0	0	0
76	0	0	0	0	0	0	0	0	0	0	0	0	0	0
77	0	0	1	0	1	0	0	0	0	0	0	0	0	0
78	1	0	0	0	0	0	0	1	0	0	0	0	0	0
79	1	0	0	0	0	0	0	1	0	0	0	0	0	0
80	1	1	0	0	0	0	0	0	0	0	0	0	0	0
81	1	0	0	0	1	1	0	0	0	0	0	0	0	0
82	1	1	0	0	0	0	0	0	0	0	0	0	0	0
83	1	0	0	0	0	1	0	0	0	0	0	0	0	1
84	0	0	0	0	0	0	0	0	0	0	0	0	0	0
85	0	0	1	0	0	0	0	0	0	0	0	0	0	0
86	1	0	0	0	0	0	0	0	0	0	0	0	0	1
87	1	1	0	0	0	0	0	0	0	0	0	0	0	0
88	0	0	0	0	0	0	0	0	0	0	0	0	0	0
89	1	0	0	0	1	0	0	0	0	0	0	0	0	0
90	1	0	0	0	0	0	0	0	0	0	0	0	0	1
91	1	1	0	0	0	0	0	0	0	0	0	0	0	0
92	1	1	0	0	0	0	0	0	0	0	0	0	0	0
93	1	0	0	0	0	0	0	0	0	0	0	0	0	1
94	1	0	0	1	0	0	0	0	0	0	0	0	0	1
95	0	0	0	0	1	1	0	0	0	0	0	0	0	1
96	1	0	1	0	1	0	0	0	0	0	0	0	0	1
97	1	0	0	0	0	0	0	0	0	0	0	0	0	1
98	0	0	0	0	0	0	0	1	0	0	0	0	0	1
99	1	0	0	0	0	1	0	0	0	0	0	0	0	1
100	1	0	1	0	0	0	0	0	0	0	0	0	0	1
101	1	0	1	0	1	0	0	0	0	0	0	0	0	1
102	0	1	0	1	0	0	0	0	0	0	0	0	0	1
103	0	0	0	0	1	1	0	0	0	0	0	0	0	1
104	1	0	0	1	0	0	0	0	0	0	0	0	0	1
105	0	0	0	0	1	0	0	0	0	0	0	0	0	1
106	1	0	1	0	0	0	0	1	0	0	0	0	0	1
107	1	0	0	0	0	1	0	0	0	0	1	0	0	1
108	1	1	0	0	1	0	0	0	0	0	0	0	0	1
109	0	1	0	0	0	1	0	0	0	0	0	1	0	0
110	0	1	0	0	1	0	0	0	0	0	0	0	0	0
111	1	0	0	0	1	0	0	0	0	0	0	0	0	0
112	1	0	0	0	0	0	0	0	0	0	0	1	1	0
113	1	0	0	0	0	0	0	1	0	0	0	0	0	0
114	1	0	0	0	1	0	0	0	0	0	0	0	0	0
115	0	0	0	0	0	0	0	0	0	0	0	0	0	0
116	0	0	0	0	1	0	0	0	0	0	0	0	0	1
117	0	0	0	0	0	0	0	0	0	0	0	0	0	0
118	1	1	0	0	1	0	0	0	0	0	0	1	0	0
119	0	0	0	0	1	0	0	0	0	0	0	0	0	1
120	1	0	1	1	1	0	0	0	0	0	0	0	0	0
121	1	0	0	0	0	0	0	0	0	0	0	0	0	1
122	1	0	0	0	1	1	0	0	0	0	0	0	0	0
123	0	0	0	0	0	0	0	0	0	0	0	0	0	0
124	1	1	0	0	0	0	0	0	0	0	0	0	0	0
125	0	0	0	0	0	0	0	0	0	0	0	0	0	0
126	1	1	0	0	0	0	0	0	0	0	0	0	0	0
127	0	0	0	0	1	1	0	0	0	0	0	0	0	1
128	0	0	0	0	1	1	0	0	0	0	0	0	0	1
129	0	0	0	0	0	0	0	0	0	0	0	0	1	0
130	1	0	0	0	0	0	0	1	0	0	0	0	0	0
131	0	0	0	0	0	1	0	0	0	0	0	0	0	1
132	0	0	0	0	0	0	0	0	0	0	0	0	0	0
133	0	0	0	0	1	1	0	0	0	0	0	0	0	1
134	0	0	0	0	0	1	0	1	0	0	0	0	0	0
135	1	0	0	0	0	0	0	1	0	0	0	0	0	0

2) Developing ANN architecture

The form of ANN architecture is based on the data of PPI-based drug regimen. Let the date as a data set S = {(x_1, y_1), (x_2, y_2), …, (x_n, y_n)}, with x_i as data vector with y_i as target class of i^{st} data. Data number 1-95 was listed as training data and data number 96-135 was listed as testing data. Data set S was taken from training data.

3) Training ANN using LVQ

Train the data of PPI-based drug regimen using LVQ2.1 algorithm. Several prepocessing to get the best of LVQ parameters using these data was done. The best parameters of LVQ were learning rate = 0.31, the width of window = 0.5, maximal training iterations = 1000, and the number of code-book vectors = 50. For each vector x_i with i = 1, 2, …, n, in accordance with LVQ2.1 training, y_i' results in output class of LVQ2.1. However, y_i' may be different from y_i.

There were 12 data which had different classify result or y_i' is not same with y_i, i.e data number: 24, 33, 58, 61, 67, 68, 71, 74, 75, 77, 81, and 89.

4) Testing the data and save as new data

By combining x_i and y_i', an instance (x_i, y_i') was generated. A data set S' = {(x_1, y_1'), (x_2, y_2'), …, (x_n, y_n')} as a data set as result of LVQ2.1 training.

5) Extracting rule using C4.5

Make an induction using the data set of training with C4.5. The results of the induction of data set S' are rules as guidelines for the PPI-based drug regimen based on the trained data. There are 18 rules and listed in Table III.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

TABLE III. THE RULES AS RESULTS OF EXTRACTION WITH LVQ-C4.5

No	IF	Then
1	Heart pain = Yes	PPI drug regimen = Yes
2	Heart pain = No and Lower abdominal pain = Yes	PPI drug regimen = No
3	Heart pain = No and Lower abdominal pain = No and Abdominal pain up to waist = Yes	PPI drug regimen = No
4	Heart pain = No and Lower abdominal pain = No and Abdominal pain up to waist = No and Stomach cramp = Yes	PPI drug regimen = Yes
5	Heart pain = No and Lower abdominal pain = No and Abdominal pain up to waist = No and Stomach cramp = No and Difficult to swallow = Yes	PPI drug regimen = Yes
6	Heart pain = No and Lower abdominal pain = No and Abdominal pain up to waist = No and Stomach cramp = No and Difficult to swallow = No and Abdominal bloating = Yes and Nausea/vomit = Yes	PPI drug regimen = Yes
7	Heart pain = No and Lower abdominal pain = No and Abdominal pain up to waist = No and Stomach cramp = No and Difficult to swallow = No and Abdominal bloating = Yes and Nausea/vomit = No	PPI drug regimen = No
8	Heart pain = No and Lower abdominal pain = No and Abdominal pain up to waist = No and Stomach cramp = No and Difficult to swallow = No and Abdominal bloating = No and swelling abdomen = Yes	PPI drug regimen = Yes
9	Heart pain = No and Lower abdominal pain = No and Abdominal pain up to waist = No and Stomach cramp = No and Difficult to swallow = No and Abdominal bloating = No and swelling abdomen = No and Fullness = Yes and Upper abdominal pain = Yes	PPI drug regimen = No
10	Heart pain = No and Lower abdominal pain = No and Abdominal pain up to waist = No and Stomach cramp = No and Difficult to swallow = No and Abdominal bloating = No and swelling abdomen = No and Fullness = Yes and Upper abdominal pain = No and Stomachache = Yes and Nausea/vomit = Yes	PPI drug regimen = No
11	Heart pain = No and Lower abdominal pain = No and Abdominal pain up to waist = No and Stomach cramp = No and Difficult to swallow = No and Abdominal bloating = No and swelling abdomen = No and Fullness = Yes and Upper abdominal pain = No and Stomachache = Yes and Nausea/vomit = No	PPI drug regimen = Yes
12	Heart pain = No and Lower abdominal pain = No and Abdominal pain up to waist = No and Stomach cramp = No and Difficult to swallow = No and Abdominal bloating = No and swelling abdomen = No and Fullness = Yes and Upper abdominal pain = No and Stomachache = No	PPI drug regimen = Yes
13	Heart pain = No and Lower abdominal pain = No and Abdominal pain up to waist = No and Stomach cramp = No and Difficult to swallow = No and Abdominal bloating = No and swelling abdomen = No and Fullness = No and Upper abdominal pain = Yes and Nausea/vomit = Yes	PPI drug regimen = No
14	Heart pain = No and Lower abdominal pain = No and Abdominal pain up to waist = No and Stomach cramp = No and Difficult to swallow = No and Abdominal bloating = No and swelling abdomen = No and Fullness = No and Upper abdominal pain = Yes and Nausea/vomit = No	PPI drug regimen = Yes
15	Heart pain = No and Lower abdominal pain = No and Abdominal pain up to waist = No and Stomach cramp = No and Difficult to swallow = No and Abdominal bloating = No and swelling abdomen = No and Fullness = No and Upper abdominal pain = No and Nausea/vomit = Yes and Stomachache = Yes	PPI drug regimen = No
16	Heart pain = No and Lower abdominal pain = No and Abdominal pain up to waist = No and Stomach cramp = No and Difficult to swallow = No and Abdominal bloating = No and swelling abdomen = No and Fullness = No and Upper abdominal pain = No and Nausea/vomit = Yes and Stomachache = No	PPI drug regimen = Yes
17	Heart pain = No and Lower abdominal pain = No and Abdominal pain up to waist = No and Stomach cramp = No and Difficult to swallow = No and Abdominal bloating = No and swelling abdomen = No and Fullness = No and Upper abdominal pain = No and Nausea/vomit = No and Stomachache = Yes	PPI drug regimen = Yes
18	Heart pain = No and Lower abdominal pain = No and Abdominal pain up to waist = No and Stomach cramp = No and Difficult to swallow = No and Abdominal bloating = No and swelling abdomen = No and Fullness = No and Upper abdominal pain = No and Nausea/vomit = No and Stomachache = No	PPI drug regimen = No

IV. RESULTS

Furthermore, the rule of PPI-based drug regimen was tested in test data, i.e. 40 data from patients given the PPI-based drugs in Baptis Hospital of Kediri from December 2009 to February 2010. Based on the test, the rule successfully given appropriate decisions on the PPI based drug regimen among 33 patients (82.5%). The appropriateness of rules of PPI-based drug regimen was higher than decisions on the PPI-based drug regimen by physicians at the period, which was only 50%.

Wrong classifying using LVQ-C4.5 rule is data number: 109, 110, 111, 114, 118, 120, and 122. However, incorrect PPI-based drugs regimen by the doctor in Baptis Hospital of Kediri on testing data is data number: 109, 110, 111, 112, 113, 114, 115, 117, 118, 120, 122, 123, 124, 125, 126, 130, 132, 134, and 135.

V. CONCLUSION

From the results of test, the number of LVQ-C4.5 rules was lesser than that of C4.5 rules. This was because LVQ made a generalization in data first. The accuracy rate of LVQ-C4.5 rules was also higher than that of C4.5 rules, although it cannot be made sure that it will always be higher in other cases. LVQ-C4.5 rules were able to identify inappropriateness in drug regimen, so that it can be made as a guideline for the PPI-based drug regimen. This can be seen in the testing section that the accuracy rate of LVQ-C4.5 rules was higher than that of the

PPI-based drug regimen that was done by physicians in Baptis Hospital of Kediri from December 2009 to February 2010.

REFERENCES

[1] A.O. Eze and F.K. Opara, "An Expert System (Autodoc) for Diagnosing Diseases and Prescribing Medication using Visual Basic.net," *Information Technology Research Journal*, vol. 2, 2012, pp. 20-24.

[2] T. Saba, S. Al-Zahrani, and A. Rehman, "Expert System for Offline Clinical Guidelines and Treatment," *Life Science Journal*, vol. 9, 2012, pp. 2639-2658.

[3] H. Bahi, and M. Sellami, "A Connectionist Expert Approach for Speech Recognition," *The International Arab Journal of Information Technology*, vol. 2, 2005, pp. 148-153.

[4] G.G. Towell and J.W. Shavlik, "Knowledge-Based Artificial Neural Networks," *Appears in Artificial Intelligence*, vol. 70, 1994, pp. 119-165.

[5] C. Alexandra, P. Cristea, and T. Okamoto, "Neural Network Knowledge Extraction," *Rev. Roumaine des Sciences Technique, Serie EE (Electrotechn. et Energ.)*, vol. 42, 1997, pp. 477-491.

[6] P.J. Lisboa, and A.F.G. Taktak, "The Use of Artificial Neural Networks in Decision Suport in Cancer," *Neural Networks*, vol. 19, 2006, pp. 408-415.

[7] A. Jouyban, S. Soltani, and K.A. Zeynali, "Solubility Prediction of Drugs in Supercritical Carbon Dioxide Using Artificial Neural Network," *Iranian Journal of Pharmaceutical Research*, vol. 6, 2007, pp. 243-250.

[8] L. Sratthaphut, and T. Phaechamud, "Prediction of Drug Dissolution and Hardness of Indomethacin Tablets using Artificial Neural Networks and Partial Least-Squares," *Research Journal of Pharmaceutical, Biological and Chemical Sciences*, 2011.

[9] S. Godara, and R. Gupta, "Neural Networks for Iris Recognition: Comparisons between LVQ and Cascade Forward Backpropagation Neural Network Model, Architectures, and Algorithm," *IOSR Journal of Engineering*, vol. 3, 2013,.

[10] L. Fausett, *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*, New Jersey: Prentice-Hall Inc., 1994.

[11] S. Hasoon, and A.Y. Jasim, "Diagnosis Windows Problem Based on Hybrid Intelligence Systems," *Journal of Engineering Science and Technology*, vol. 8, 2013, pp. 566-577.

[12] K. McGarry, S. Wermter, and Maclntyre, "Hybrid Neural Systems : From Simple Coupling to Fully Integrated Neural Networks," *Neural Computing Surveys*, vol. 2, 1999, pp. 62-93.

[13] S.M., Kamruzzaman and M.M. Islam, "Extraction of Symbolic Rules from ANN," *arXiv preprint arXiv*: 1009.4570, 2010.

[14] T. Kohonen, *Self-Organizing Maps*, New York: Springer-Verlag Berlin Heidelberg, 3rd ed., 2008.

[15] D.T. Larose, *Discovering Knowledge in Data: an Introduction to Data Mining*, John Wiley and Sons, 2014.

AUTHORS PROFILE

Anifuddin Azis, S.Si., M.Kom. He is lecturer in Department of Computer Science and Electronics in Gadjah Mada University. He is doing doctoral degree in Computer Science in Gadjah Mada University. His research interests are Neural Network, Expert Systems, and Decision Support System.

Prof. Dra Sri Hartati M.Sc., Ph.D. She is professor in Department of Computer Science and Electronics in Gadjah Mada University. She recieved her Ph.D from Univerity of New Brunswick, Canada in 1996. Her research interest is Intelligent Systems, i.e Knowledge Based System, Reasoning System, Expert System, Fuzzy System, Pattern Recognation, Decision Support System, Medical Computing and Soft Computation.

Edi Winarko, M.Sc., Ph.D. He is senior lecturer and head of Post Graduated Program in Department of Computer Science and Electronics in Gadjah Mada University. He received his Ph.D in Computer Science from Flinders University, Australia. His research interests are Machine Learning, Data Mining, Information Retrieval.

Prof. Dra. Zullies Ikawati Ph.D. She is professor in Faculty of Pharmacy in Gadjah Mada University. She recieved her Ehime University, Japan in 2001. Her research interest is Pharmacology, Clinical Pharmacy, Pharmacogenetic/genomic.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 6, June 2016

APMS: CONSTRUCTION AND ASSESSMENT OF HOSPITAL PROCESS FOR OUTPATIENTS PROCESS ANALYSIS

Mohammad Taha Khan, *Research Scholar, Suresh Gyan Vihar University Jaipur Rajasthan India,*

Dr. Shamimul Qamar, *Professor College of Computer Science, King Khalid University Abha KSA,*

Dr. Ripu Ranjan Sinha, *Associate Dean Research, Suresh Gyan Vihar University Jaipur India,*

Abstract- **Management Information Systems is the process of transforming the accumulated data into useful and helpful information systems. This paper work is on design and construction of Advanced Pathology Management System (APMS). The objectives of the APMS is to i) Well-secured login system ii) Simple and easy patient registration form iii) Better test processing system i.e scheduling for the test and tracking the reports iv) Efficient Report Management system i.e, creation, searching and verification of the required reports v) Well-defined privacy management systems. The developed APMS is tested over Urgent care hospital, New Delhi. The event logs of outpatients are accumulated from the hospital and preprocessed using process mining approaches. Performance indices such as wait time for consultation wait time for test and the aggregate time spent on the outpatient care are analyzed. Experimental results prove the efficiency of the developed Advanced Pathology Management System (APMS).**

Keywords: Management Information Systems, Clinical Pathology, Report Management, Outpatients and Process mining approaches.

I. INTRODUCTION

Accumulation of the data will bring change, across all the fields, at a vivid pace. These data should properly utilize to extract the knowledgeable data. The most widely adopted step in Knowledge Discovery in database (KDD) is the Data Mining. The process of dealing with high data volume is known as data mining. A clear form of data is extracted from colossal of data is known as Information Extraction (IE) systems (Lenz and Richertz, 2007). The core process involved in data mining is the Pattern Discovery. In view of economic environment, data plays vital role.

Data mining acts as a baseline for the domains like machine learning, artificial intelligence, probability and statistics.

Though, there are several applications supported by the data mining. Healthcare system is the most interesting application study in data mining.

Both the healthcare industry and data mining fused together to create a great revolution to make better clinical decisions by general practitioners Mulyar, Pesic, van der Aalst, and Peleg ,2008).

In order to provide a good service to any kind of environment, an efficient and effective data mining process is involved. The cerebration of providing an unbeatable hospitality services relied upon the suitable data mining process (Maggi, Mooij, & van der Aalst, 2011). The target of healthcare processes are a chain of activities that comprised of diagnosing, treat and prevent any sort of diseases in accord to intensify their patient's health. The suggestions are provided by various resources such as physicians, nurses, expert's advice and managers. The data mining processes might vary from organization to organization. So, it's important to design an efficient healthcare processes. The well- equipped healthcare processes enlighten the life quality of the patients. Anyhow, Enlightenments in healthcare processes is not an easy task. Several challenges have to solve proficiently. The main challenge is the reduced level of service towards cost, response towards the patient's queries. Simultaneously, the resource productivity and higher transparency is also enhanced between patients and general practitioners.

In the view of medical treatment process, the process in healthcare is differentiated into two concerns- a) Therapeutic activities and b) Administrations activities (Lenz and Reichert, 2007). The administration process should be adaptable to the hospitable environment. It belongs to the class of Evidence based reasoning systems. Based on the evidences, the actions are determined and processed (Eddy, 2005). Medical guidelines are framed to formalize the medical activities.

If the guidelines are not handled properly, then it is treated an interaction gap between the clinical practices and the recommendation systems (Cochrane et al., 2007; Hay et al., 2008; Lew & DeMaria, 2013). The maintenance of the recommendation systems is a vital process in the healthcare management environment (Hetlevik, Holmen, Krger, & Holen, 1997; Milchak, Carte, James, & Ardery, 2004).

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

This is the main kingdom of process mining that introduced little decades ago, known as Clinical Systems Pathology (van der Aalst, 2011).

1.1 System pathology – overview:

System Pathology stated as "the interwoven between the functional level and morphological information into a solitary coherent model in order to percepts physiological systems and clinical pathologies" (Saidi et al., 2007). The tremendous growth in the field of computational biology activates to develop the system model using empirical data. In the ecological environment, a huge dynamics data are accumulated. These sorts of heterogeneous data are consolidated into solitary data interaction systems. Top-down and bottom-up approaches are performed to portray a system that how well the system behavior is generated. With top-down approach as base layer, the bottom-up approach is used as interaction system. It is known as 'Causality' in system pathology (Tarafa et al, 2008). Pathology is the study of causes and effects of disease. Clinical Systems Pathology is a model that invents the solution to the problem of intended patients via tools of experimental science to the clinical problem (Saidi et al., 2007; Donovan et al., 2009a; Faratian et al., 2009). Insofar in biomedical sciences, the art of developing a communication model between observation science and experimental science is of greater important towards the "Process Mining in Healthcare Systems".

Though there are more and more cases in literature discusses about the relationship between system pathology and clinical pathology, yet some descriptive challenges are to be studied. In this paper, we make an attempt to study about the significance of clinical system pathology using Open Source Technologies via Data Mining approaches. The main contributions in this paper are:

a) Discussing about the descriptive challenges pertains in the domain of Clinical Systems Pathology.
b) Discussing about the roles and responsibilities of data mining process towards clinical systems pathology.
c) Discussing about how the Open Source Technologies assisting the Clinical Systems Pathology.
d) At last, discussing about the futuristic research directions.

The paper is structured as follows: Section 2 motivates the work. The methodology followed in the case study is described in Section 3. Section 4 shows how recently a developed open source technology supports the methodology. Section 5 illustrates the experimental results. Section 6 concludes the paper.

II. MOTIVATION AND PROBLEM STATEMENT

Most of the real time applications were ruled by the Business Intelligence Process (BIP). A divergence will exist between predicted traits and the original traits. Consider a medical diagnosis processes, sometimes the clinical decisions may change depends upon the patient's condition. The main involvement in this paper is to find the root cause for this divergence and analysing it through a case study. We assure you that the divergence analysis can assist for future process implementation if they update the clinical systems model.

2.1 Problem Statement

In healthcare environment, Management Division (Bansal, Bertels, Ewart, MacConnachie, & O'Brien, 2012; Cahill & Heyland, 2010; Dresselhaus, Peabody, Lee, Wang, & Luck, 2000; Hay et al., 2008) is the main part to deal with the divergence between clinical directions and the process deployments. The factors involved in the clinical directions are reviewed and depicted as follows:

a) Clinical Prescriptions illustrates the divergence in the actual traits. A structural interview can also influence the clinical decisions (Freedman and Sweney, 2001). .
b) Cost: The patient's economic background doesn't suits the cost of treatments (Bernheim, Ross, Krumholz, & Bradley, 2008).
c) The synchronization in several pathologies (e.g. diseases like allergies, surgeries) has to be improved. This kind of situation can prolong the patient's treatment process.
d) Some research results are not widely used in practice (Graham et al, 2006).
e) Data privacy preserving is also becoming an eminent issue in healthcare sectors (Berenholtz and Pronovost, 2007).

The demerits in the existing Clinical Systems Pathology are:

Firstly, the divergence is studied by conducting interviews; discussion with patients and experts, Observational studies and audits (Page et al, 2010; Hajjaj et al., 2010; Lew & DeMaria, 2013). The analysis is manual that leads to high time consuming and error-prone.

Second, with the advent of IT exploration systems, the errors are analyzed and solved by the continuous monitoring of the system activities. Some business intelligence tools were developed to monitor the activities of the system (Fichman et al, 2011; Lenz, 2007).

Third, the processing techniques involved in the data mining can't assure the scalability and reliability of the information. Relied upon the previous activities, the medical errors are rectified and the new service is developed (Rebuge & Fer-reira, 2012; Weerdt, Caron, Vanthienen, and Baesens, 2013).

III. RESEARCH METHODOLOGY

3.1 Materials

The study site, Urgent Care hospital, New Delhi, is a tertiary hospital located in New Delhi that comprised of provisions listed in Table 3.1

List of Provisions	Totals (in numbers)
Beds	110
Medical departments	5
Diagnostic laboratories	6
Surgical departments	2
Operating rooms	4
Nursing wards	16
Central Pharmacy	1
Central material departments	1

Table 1: List of Provisions at Urgent Care hospital.

The Urgent care hospital is established in 1990's. This widely spread across multiple centers. It is well-known for its medical services towards emergency of the patients. They serve the people with efficient hospitality service at the reduced monetary level. The emergency services handled are the: 24 * 7 Ambulance care, Asthma, ECG, Digital X-Rays, USG, Cardiac monitoring, Thrombolysis, Ventilators, Nebulizers, Cardiac, Pulmonary flow rates, Laboratories, Pharmacy, Urinary Catheterization, Ear Wax Removal, Feeding tube, and Injection Administration. The objectives of the study are:

- To effectively allocate the bed allocation.
- To effectively use the operation theatres.
- To analyse how the emergencies situation affects the Administration system.
- To provide better pathology decisions.
- To find the cluster of patients.

The above mentioned objectives related to the emergency case where the hospitality service can't be delayed. Anyhow, all the objectives are interrelated to each other. For an instance, consider a surgery is frequently cancelled; this will significantly impact the patients in the waiting list.

Presently, these clinical objectives are transformed into Objectives of the data mining.

- To develop a data mining model that effectively handles in-patients and out-patients of emergencies into variant time periods i.e. Works shifts, daily activities.
- To provide better hospitalization services.
- To develop predictive model to processes the request.
- To develop a model how the hospital resources influences the diseases.
- To carry out models to cluster patients (by age, by area, by pathology class, etc).

Hypothesis Settings:

H_0 – There is no serious change to the hospital administration services in any emergency scenarios using Open Source Technologies.

H_1- There is a serious change to the hospital administration services in any emergency scenarios using Open Source Technologies.

a) Event Log collection and Preprocessing:

The event log of the outpatient care process, collected from Urgent Care Hospital, is listed in Table 2. The information like Task completion time, department and their related information are collected when the patient visits their hospitals.

The attribute, Case ID is included in order to distinct the events according to the patients.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Task (event type)	***Attribute***
Selective medical service	Case ID, Activity completion time, Resource ID, Resource department code
Registration for referrals	Case ID, Activity completion time, Resource ID, Resource department code
Outside image analysis	Case ID, Activity completion time, Resource ID, Resource department code
Payment	Case ID, Activity completion time, Resource ID, Resource department code
Test registration	Case ID, Activity completion time, Resource ID, Resource department code, Test code, Type of test, Scheduled test date
Lab Test Generation	Case ID, Activity completion time, Resource ID, Resource department code, Test code, Type of test, Scheduled test date
Registration for consultation	Case ID, Activity completion time, Resource ID, Resource department code, Patient type, Department code, Appointment method, Appointment Date
Consultation	Case ID, Activity completion time, Resource ID, Resource department code, Patient type, Department code, Appointment method, Appointment Date
Scheduling for consultation	Case ID, Activity completion time, Resource ID, Resource department code, Patient type, Practitioner ID, Scheduled department code, Scheduled consultation date
Scheduling for test analysis	Case ID, Activity completion time, Resource ID, Resource department code, Test code, Type of test, Scheduled test date
Scheduling for admission	Case ID, Activity completion time, Resource ID, Resource department code
Certificate issuing	Case ID, Activity completion time, Resource ID, Resource department code

Table 2: Task and attributes of event logs

b) Process mining technique via Open Source Technologies:

The investigation of business process requirements based on the acquired event logs is known as process mining. The data obtained from Urgent care hospital is examined according to frequency analysis of the case, the hourly distribution of patients, aggregate time of outpatient care, time spent per task and patient to task ratio.

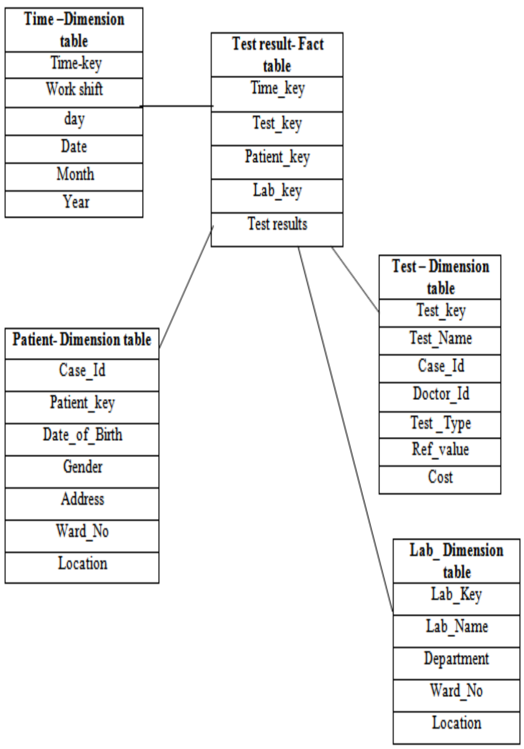


Fig.1. Creating fact tables and dimension tables for Urgent Care hospital

Based on the attributes listed in Table 2, we develop Advanced Pathology Management System (APMS) using Open Source Technology, Hypertext Preprocessor (PHP). PHP is an eminent scripting language that aims to enhance the visual aspects of the websites. Here, it is widely used in the field of Clinical Pathology Systems. The proposed Advanced Pathology Management system (APMS) covers the following requirements:

a) Well-secured login system
b) Simple and easy patient registration form
c) Better test processing system i.e scheduling for the test and tracking the reports.
d) Efficient Report Management system i.e, creation, searching and verification of the required reports.
e) Well-defined privacy management systems.

IV. EXPERIMENTAL RESULTS

In this section, we justify the efficiency of the Advanced Pathology Management System (APMS). The sample screenshots are given below:

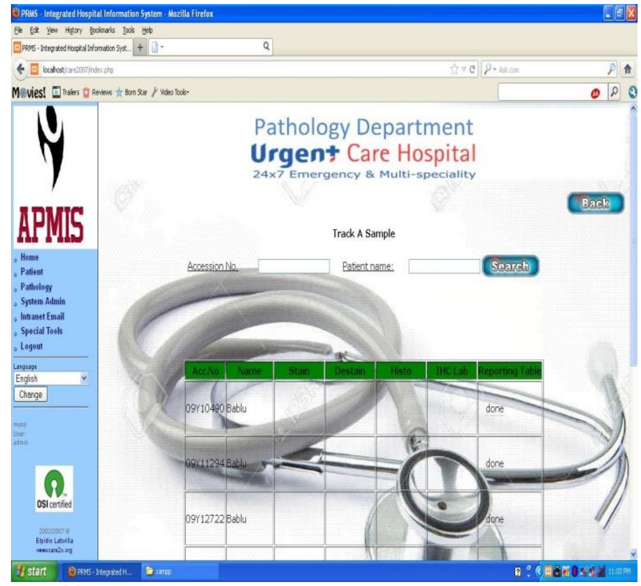


Fig.2. Sample tracking the reports of a patient.

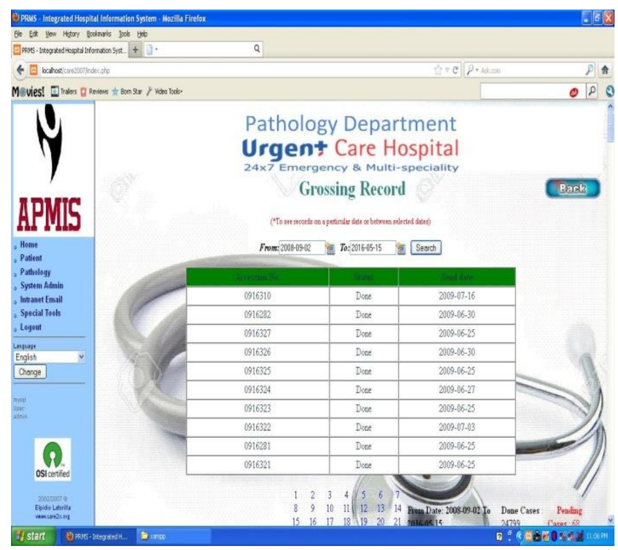


Fig.3. View the statistics

In order to assess the efficiency of the outpatient care, the performance indices such as wait time for consultation, wait time for test and the aggregate time spent on the outpatient care w.r.t changes in number of patients.

The time consumption occur between the patient visits the hospital till the treatment end is known as aggregate time spent on the outpatient care. The wait time for consultation is the period of time slot when the patient gets registered until the consultation begins. The wait time for test is the period of time slot after the test is registered until the time when the test begins. Consider any two departments viz, Urology and Cardiology center. The performance indices are measured by p- value frequency analysis. P-value is the level of significance of the events towards the null hypothesis.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

	Urology center		Cardiology center	
	Total values	p-values	Total values	p-values
Total no.of outpatients	1000		1340	
Aggregate time for outpatient care	116.95	0.025	88.24	0.560
Wait time for consultation	23.35	0.275	27.15	0.005
Wait time for test	10.96	0.006	7.19	0.115
Patient to task ratio	0.58	0.039	0.85	0.189

Table 3: Obtaining p-values

V. RELATED WORK

In the business process perspectives, hospitals are regarded as Enterprises in relevant to high technology and Information systems. These types of organization are not hierarchically structured but efficient in handling better decision process (Lawrence and Dyer, 1982; W.M.P. Van der Aalst et al, 2007). When a survey conducted in European hospital, it found that the technologies can also influence the hospital events and services (Anderson, 1993; M. Song et al, 2013). Later the advent of IT systems will also offers a great support to the hospital management systems (Smith, 1999; W. M. P. Van der Aalst et al, 2004, 20011).

Healthcare management is the growing field with lot of opportunistic challenges in both direct and non-direct care scenarios (Thompson, 2010, R.M. Werner, 2010). Analysis over the event logs can also enhance the healthcare process to take better clinical decisions. Simultaneously, it intensifies the operational efficiency of the healthcare management process. The establishment of Electronic Medical Records (EMR) and the exploration of this record pave a way to enhance the degree of patient satisfaction, enhancing hospital efficiency and healthcare quality, protecting the safety of healthcare, and reducing healthcare costs (R. Mans et al, 2009; E. Kim et al, 2013).

Several different user groups like physicians, nurses, administrators, managers, radiologists, pharmacists, etc with variety of backgrounds exist in healthcare organizations (R. Mans, 2015). Implementation of a hospital information system could not happen without an analysis of the perceptions of patient's satisfaction that make use of Hospital Management system (Ndira, Rosenberger, and Wetter, 2008).

VI. CONCLUSION

The quality of hospital services depend upon the suitable provision and well-defined processing systems. Healthcare process is a chain of tasks carried out to diagnose, treat and prevent the patients from diseases. The aim of the healthcare process is to enhance the patient's health with lessened cost and high quality services. In this paper, we develop "Advanced Pathology Management Systems" that target to overcome the challenges in Urgent Care Hospital, New Delhi. The outpatient's event logs are accumulated and preprocessed via process mining approach. The wait time of the outpatient's treatment is analyzed and measured using frequency analysis approach. At last, we concludes that the there is no significant changes to the hospital administration services in any emergency scenarios using Open Source Technologies. As a future work, the analysis of inpatient's care will be studied.

REFERENCES

[1] W.M.P. Van der Aalst, A.J.M.M. Weijters, B.F. Van Dungeon, A.K. Alves de Medeiros, M. Song, H.M.W. Verbeek, Business process mining: an industrial application, Inf. Syst. 32 (2007). 713–732.

[2] M. Song, S.H. Siadat, M. Pechenizkiy, A comparative study of dimensionality reduction techniques to enhance trace clustering performances, Exp. Syst. Appl. 40 (2013). 3722–3737.

[3] W.M.P. Van der Aalst, J. Herbst, L. Maruster, G. Schimm, A.J.M.M. Weijters, Workflow mining: a survey of issues and approaches, Data Knowl. Eng. 47 (2003). 237–267.

[4] W.M.P. van der Aalst, M.H. Schonenberg, M. Song, Time prediction based on process mining, Inf. Syst. 36 (2011) 450–475. [5] W.M.P. Van der Aalst, A.J.M.M. Weijters, L. Maruster, Workflow mining: discovering process models from event logs, IEEE Trans. Knowl. Data Eng. 16 (2004). 1128–1142.

[6] C.W. Gunther, W.M.P. Van der Aalst, Fuzzy mining: adaptive process simplification based on multi-perspective metrics, Proceedings of the 5th International Conference on Business Process Management (2007) 328–343.

[7] J.A. Linder, J. Ma, D.W. Bates, B. Middleton, R.S. Stafford, Electronic health record use and the quality of ambulatory care in the United States, Arch. Intern. Med. 167 (2007) 1400–1405.

[8] R. Mans, H. Schonenberg, G. Leonardi, S. Panzarasa, A. Cavallini, W.M.P. van der Aalst, Process mining techniques: an application to stroke care, Stud. Health Technol. Inform. 136 (2008) 573–578.

[9] E.D. Peterson, M.T. Roe, J. Mulgund, E.R. DeLong, B.L. Lytie, R.G. Brindis, S.C. Smith Jr., C.V. Pollack Jr., L.K. Newby, R.A. Harrington, W.B. Gibler, E.M. Ohman, Association between hospital process performance and outcomes among patients with acute coronary syndromes, JAMA 295 (2006) 1912–1920.

[10] R.M. Werner, E.T. Bradlow, Public reporting on hospital process improvements is linked to better patient outcomes, Health Aff. 29 (2010) 1319–1324.

[11] R. Mans, M.H. Schonenberg, M. Song, W.M.P. van der Aalst, P.J. Bakker, Application of process mining in healthcare: a case study in a Dutch hospital, in: A. Fred, J. Filipe, H. Gamboa (Eds.), Biomedical Engineering Systems and Technologies, Springer, Heidelberg, 2009, pp. 425–438.

[12] P. Gemmel, D. Vandaele, W. Tambeur, Hospital Process Orientation (HPO): the development of a measurement tool, Total Qual. Manage. 19 (2008) 1207–1217.

[13] E. Kim, S. Kim, M. Song, S. Kim, D. Yoo, H. Hwang, S. Yoo, Discovery of outpatient care process of a tertiary university hospital using process mining, Healthcare Inform. Res. 19 (2013) 42–49.

[14] W.M.P. Van der Aalst, A.J.M.M. Weijters, A.K. Alves de Medeiros, Process mining with the heuristics miner-algorithm, Technische Universiteit Eindhoven, Tech. Rep. WP 166 (2006) 1–34.

[15] M. Song, W.M.P. van der Aalst, Towards comprehensive support for organizational mining, Decis. Support Syst. 46 (2008) 300–317.

ABOUT AUTHORS

Mr. Mohammad Taha Khan is a Research Scholar at Suresh Gyan Vihar University Jaipur. He has B.E (electronics) from Jiwaji University Gwalior and M-Tech (I.T) from C-DAC Noida, His research interest includes e-Health care, Open Source Systems and data mining.

Dr. Shamimul Qamar has about 17 years experience in academics and research.He has M.Tech from AMU, Aligarh and Ph.D. in Computer Science & Engg from IIT Roorkee India. Presently, he is working as Professor at King Khalid University , Abha KSA . His research interest include Artificial Intelligence ,Open Source Systems , wireless and mobile networks .

Dr. Ripu Ranjan Sinha, PhD (CS), with one decade of Experience in IT Training and Administration at AICTE Approved institution. Currently he is working as Associate Dean Research with Suresh Gyan Vihar University. His research interests include Knowledge Management, Enterprise resource Planning.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Anonymity of Base Station in Wireless Sensor Network via Backup Base Station

Faisal Taj[1], Shahzad Anwar[2*], Muhammad Imran[1], and Irshad Ullah[1]
[1]Department of Computer Science, Iqra National University, Peshawar Pakistan
[2]Institute of Mechatronics, University of Engineering and Technology Peshawar, Pakistan

Abstract - Sensor nodes covers surrounding area and report any events to a base station over multi-hop communication. The base station plays a key role in the network. The adversary, wants to disrupt network operation, would excitedly look for the base station and target it with attacks in order to inflict maximum damage. To avoid maximum damage a novel approach is proposed for boosting the anonymity of the base station. In the proposed research the numbers of base stations are increased from one to many (such as 2 to 5) in the network operation. The purpose is to divert the adversary attention about the base station and adversary considers the base station as a sensor node. Experimentation results suggest that the approach provide a backup facility in case if one of the base stations is failed due to adversary or due to energy failure. Therefore enhances network security.

Keywords –Anonymity, Base Station, Backup Base Station, Wireless Sensor Network

I. Introduction

Wireless Sensor Network (WSN) have many application such as, military battle field surveillance [1], environmental disasters e.g. forest fire detection [2] or in sensitive health applications [3] etc. The use of WSN in military is to monitor enemy activities and to ensure force protection in remote areas. The network equipped with suitable sensor enables opponent's movement detection, and study their behavior [4]. The sensor nodes (in a network) collect data from the environment [i.e. direct coverage area] and send it to base station for processing. The central point of WSN which is the Base Station (BS) is responsible to handle and control the whole network. Data collection by the sensor nodes in the direct coverage area is transmitted to Base Station (BS) by using the nodes in BS coverage area. So the transmission of data from direct coverage area to Base Station is via multi-hop communication. The Base Station passes on the information received from the sensor nodes to command center. The command center is the place where this information is required for area monitoring purposes.

In a network, flow of traffic [i.e. transmission of data packets] from the sensor node to the BS is fixed [5], and this behavior of the network helps adversaries to detect BS location. The adversaries are the forces from opponent that attempt to disrupt the network by performing traffic analysis [6]. The transmission paths near BS are usually condensed /merged and the presence of data packets near BS is comparatively high [i.e. high rate of data traffic near BS]. These also increase the chances for BS detection and resultantly increases the vulnerability for an eminent attached from adversaries. Adversaries become closer to the BS by data traffic scanning and once BS detected, destroy the BS making the whole network useless [7].

The energy source is also a major issue in WSN and can be a cause for BS failure. The lifetime of sensor node is dependent on battery power [i.e. energy source] which in turn is dependent on energy consumption [8]. The rate of consumption is directly proportional to communication distance and is higher for long distance communications. It is worth mentioning that the multi-hop communication between sensor nodes and BS is popular in large scale wireless sensor network than single hop communication [9]. In multi-hop communication the sensor nodes spend most of its energy on routing data packets so the consumption rate is higher and effects battery power. Therefore it is important to shorten the hop distance between each sensor node and the base station [10] to save energy consumption [11].

To address these limitations a novel approach has been proposed that offers BS protect from a possible attack of an adversary and serve as back-up too if BS is destroyed from battery power failure [energy failure] or from an adversary attack. The approach will employ two base stations at a hidden location. One base station will serve as a Back-up Base Station [herein after referred as BBS] whereas the second base station [herein after referred as BS] will work for routine functions.

The rest of the paper is organized as following. Initially, Literature review is and related work is described in Section II. Following, section III presents the problem statement. System modeling and simulation is discussed in section IV. Results are presented in section V. Finally section VI conclude the paper.

II. RELATED REVIEW

Acharya et.al [12] has proposed two different approaches for anonymity of the base station [BS].

1. One approach is: BS dynamically changes its position within the network and all the traffic is diverted to the new BS position. The change in position of BS enhances the anonymity and its impact on the network performance.

2. Second is that BS behaves like a sensor node and resend

data packets which BS has received from other sensor nodes. The purpose is to confuse adversary for not considering it a data sink. Both approaches can protect base station [BS] from adversaries only.

To attain all anonymities Juan et.al [14] proposed an efficient anonymous communication (EAC) protocol. EAC consist of four well-organized anonymous scheme for data sending, forwarding, broad casting and acknowledgement (ACK). EAC uses hashing function and symmetric cryptography. As compare to existing anonymous communication protocol EAC provides complete anonymity though incurring small storage computation & communication cost. Although Destination Controlled Anonymous Routing Protocol for sensor network (DCARPS) has the lowest storage and computation cost but DCARPS has the worst anonymity & security performance. They cannot accomplish the base station and communication relationship anonymity under global passive attacks; they also can't defend the active attacks. Similarly in the case of simple anonymity scheme (SAS). The EAC Protocol provides all three types of anonymities i.e. sender anonymity, communication relationship anonymity and base station anonymity and also has low overhead.

Alomair et.al [15] proposed a new model for evaluating and analyzing the anonymity in sensor network. The novelty of the model is that it also detect a source of "information leakage" that cannot be determined using existing models. To analyze the anonymity in WSN the author introduce a new notion of "interval indistinguishability" which is stronger than existing notions and capture the source of "information leakage" that is untraceable by existing notions. The quantities measure is used to evaluate the anonymity in WSN. The objective of this work is not to propose a specific design for anonymous systems but to provide a general, security oriented, model for evaluation & analyzing the security of anonymous systems.

In many application of WSN the data privacy it may not be as important as the source location privacy. Besides the privacy of source location, BS location privacy should also be provided. Due to the open nature of WSN end-to-end privacy solution would be a tough task to attain. The anonymity and observability are the key schemes needed for end-to-end location privacy. Abuzneid et al introduce a framework called Fortified Anonymous communication (FAC) protocol for WSN. FAC protected against a complicated threat model. This work also give a solution of end to end anonymity and privacy of location [13].

To counter traffic analysis and boosting the anonymity Ren et.al [16] proposed three approaches. In the first approach the effect of multiple BS is studied, the addition of more BS increase the performance since the traffic will be spread, the collected data of sensor node is divided into multiple BS. Since the BS is much more expensive than sensors, a performance and cost trade-off exists. The second idea of this research is to utilize the mobility of BS and the effect of few BS to relocate itself into lowest anonymity regions can be categorized. The other Ren et al proposed method is to group the sensors into clusters which are managed by cluster head. The sensors in the cluster collects data analyze the data and forward over multi-hop paths to the cluster-head. The technique is known as dynamic re-association algorithm.

The WSN are vulnerable to a huge number of Security threats like privacy, confidentiality, availability, integrity and authentication. Ranjani et.al [17] briefly introduce these types of attacks and surveyed the location privacy in the WSN. Location privacy in the WSN can be categories into "content-oriented privacy" and "context-oriented privacy". The location and traffic transmission timing is focused in context oriented privacy. The special sensor nodes such as BS and Data source need location privacy, the opponent with the information of data source and BS location may be able to gather the content of the transmitted data or destroy the network. In timing privacy, An adversary concerns the time when data source collect the data and forward to the BS over multi-hope path.

Badry et.al [18] Proposed Multi-player Anonymity optimization Game (MAG) for location anonymity in a multi Base station WSN. The basic idea is to complicate the traffic analysis for boosting the anonymity of BSs in Wireless ad-hoc networks. The purpose is to confuse the opponent and avert its concentration away from the BS location. A game is formulated to dynamically handle the traffic pattern in the existence of multiple BS. Each BS selectively forwards a portion of its traffic to other BS to increase its location anonymity. The target BS is selected in a way that the variance in the "location anonymity" over all BS is minimized.

Edith et.al [19] Provide sink anonymity for sensor network in which the sensor node is unaware about the nod ID and location when routing the message. The proposed technique is known as randomized routing with hidden address (RRAH).In the proposed technique the sensor node do not know the BS and its location when transmitting the packet because the packet header do not contain the destination field. The messages are transmitted from source to the BS along a random path without a particular destination. The packet continuously traversing until a predefined hop count is reached. When the transmitted packet arrives the BS, the BS will decrypt the packet and read it silently. The disadvantage of this method is that the hop count assigned to each message has to be large to ensure that the message reaches the sink node before dying.

Li et.al [20] propose a scheme in which Anonymous Topology Discovery (ATD) and intelligent fake packet injection (IFPI) are used to secure the confidentiality of location of BS. ATD remove the possible threats for the BS during topology detection phase. In Data transmission phase IFPI improve the confidentiality of location protection.

The consequences of network synchronization on BS anonymity are not to be considered during the evaluation of proposed traffic analysis techniques. The evaluation is to be done on the basis of data traffic. Ward et.al [21] Use the Evidence Theory (ET) to observe the effect of Synchronization on anonymity performance. For this purpose Reference Broadcast Synchronization (RBS) and Timing-synch Protocol for Sensor Network are to be considered.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Which also exhibit that the appropriate configuration of Synchronization protocol boosts the anonymity of BS without consuming extra energy.

Conner et.al [22] design a technique to prevent the traffic analysis and reduce the power consumption in WSN. The basic idea is that the sensor node in the vicinity collects data and send to the decoy sink node. The decoy sink node is a temporary BS forward the aggregated data to the real BS. The technique combining indirection and data aggregation scheme, all the traffic generated away from the original BS to create traffic analysis more complex for the adversary. The disadvantage of the technique is that if the adversary find out the location of the decoy sink node they will destroy it and the entire network is useless. The adversary in the WSN capture the radio signals from the sensor node and implement the traffic analysis scheme to find the location of the BS. To harder the traffic analysis for the adversary and boost the anonymity of BS Ebrahimi et.al [23] Proposed a novel approach to enhance the transmission power of the sensor node. The increase in transmission power complex the traffic analysis done by the opponent and the level of ambiguity about the BS location raise.

In the proposed approach the BS will resend data packets only towards BBS which will reduce congestion within network. In case if BS is destroyed, BBS will take over the charge and will continue the operations of network smoothly. Although the additional BBS in the network will be an extra cost both in term of money as well as in term of routing traffic. But it will massively decrease the downtime of network and in-fact will continue the operations of the network smoothly. Moreover the data will remain secure as well due to back-up facility.

III. PROBLEM STATEMENT

It is challenging to protect BS in a network from various attacks; therefore, this research offers a backup facility after a successful attack on BS from the adversary.

IV. SYSTEM MODEL AND SIMULATION

We consider a wireless sensor network containing a Base Station, Backup Base station and a large number of sensor nodes. All nodes have the same radio range. The network work in three phases, in first phase each node finds its location and informed the Base station about its location. In the second phase the sensor node and BBS collects the data from the environment and send it to the BS. In third phase BBS send a message to the BS, in response BS send collected data to the BBS (which also ensure the life status of both base stations). If the BS is failed to reply then BBS take over the charge of BS and operate the network smoothly.

Alomair et.al [15], and Shao et.al [24] defined the adversary as external, passive and global. We consider global adversary because all the communication between the sensor node are monitored and analyze continuously. The signal strength and time of every packet transfer from the source to the destination is observed. The adversary used proxy antenna in order to reveal the network architecture.

As earlier described in [12] the objective of the adversary was to provide maximum damage by utilizing minimum efforts. The adversary can spy the whole network by using their proxy antenna to intercept the transmitted traffic. The motivated adversary has the ability to place their own network to capture the radio signals in the vicinity. The adversary gets the individual node information by using the localization technique. The adversary only detect and capture the packet, but does not interest to decrypt the content of the packet. The adversary avoids injecting its packet in the network and supposing to take enough precautions to avoid detection.

The adversary consider in our research work is Non-malicious which means the adversary does not interfere in the network communication. The only aim of the adversary is to get the location of the BS. The adversary may have a strong and powerful hardware like unlimited energy, memory and computation capabilities. It has the ability to find the location of the destination through the signal strength of the BS. The adversary is unaware about the content of the packets. The adversary only observes the delivery path but not entire path within its range because the adversary moves in the network much slower than the packet.

The adversary wants to disrupt the overall network would not be interested to find the location of the BS. Adversary just wants to find the area of the network in which the BS lies. The adversary divide the entire network into a grid of equal size cells and use their proxy antennas at cell level, the adversary assume that the cell is a Surveillance Area. He reduces the size of the set of anonymity by excluding those cells which has low confidence that the BS is present. Adversary would endeavor to identify the surveillance area where the BS is more likely to present. He simultaneously monitors the radio transmission by their signal detection antennas and gathers information by using their anonymity models. The models for measuring the anonymity are defined as following.

Entropy is a metric used to measure the anonymity of the BS. The adversaries who want to find the location of the BS divide the whole network into a grid of equal size cells. A probability is assign to each cell that the BS is lies in the cell boundary. At the start each cell has the equal probability. If the total number of cell is denoted by N then the probability of each cell is 1/N. In an ideal case the grade of anonymity is 1 and it decreases with the passage of time because of system information leak or system gives hints. When the adversary successfully finds the BS the cell probability becomes 1 and grade of anonymity is (Zero) 0. This means that the anonymity value range is between 0 and 1.

The metric GSAT score is used to measure the anonymity of the BS. The GSAT test assumes a local eavesdropper and start at a random position to analyze the traffic in the vicinity. The GSAT follow a greedy approach to find the BS. It moves to the region with higher transmission hoping to find the BS. If the adversary engaged in local maxima, which do not contain BS moves in random directions, and continues the search on the basis of observe transmission. The process is continuing until the BS is found. The total number of hopes that took to

 https://sites.google.com/site/ijcsis/
ISSN 1947-5500

reach BS is the GSAT score. The greater the number of hopes the greater adversary take time to locate BS.

To address above limitations a novel approach has been proposed that will not only protect BS from a possible attack of an adversary but will also work as back-up if BS is destroyed from battery power failure [energy failure] or from an adversary attack. The approach will employ two base stations at a hidden location. One base station will serve as a Back-up Base Station whereas the second base station will work for routine functions. Both base stations will have the ability to behave like sensor nodes by transmitting signals. BS will transmit “interest” to sensor nodes in direct coverage area for collection of data from environment. The sensor nodes will collect data from environment and transmit it to the BS. BS will process the data and at the same time will send it backward towards BBS after regular intervals. BBS not only receive data from BS but also from other sensor nodes. The data collected from other sensor nodes by BBS is sent to BS. BBS will also send a message to BS on regular intervals and in response to the message BS will send data towards BBS.

The two way communication will symbolize the live status of both base stations. Since there is no end point in the network therefore adversary will not be able to locate base station. Even if adversary finds the location of base station and destroy it or if the base station is destroyed by energy failure [battery power failure], the BBS will take charge and will continue the operation of network smoothly. Resultantly the downtime of network will also be minimized or negligible.

BBA approach has been evaluated through NS2 simulation experiments. The experiment considers has a total range of 200 to 600 sensor nodes. The number of sensor in a cell is varying from 3.33 to 30 sensors per cell. Nodes are randomly deployed in *1000* × *1000* meter area to monitor the environment and inform the BS about the target. The BBS also take a part to send the target information to the BS. The total simulation time to evaluate the performance of proposed scheme is 10 Hours. We focus on adversary efforts that use Gsat and Entropy matrices to find the location of the BS, so the experiment is taken on three different configurations of equal-sized cells, 20 Cells of Size 223 × 223m, 40 Cells of Size 158 × 158m and 60 cells of 129 × 129m. The distribution of region into cells of different sizes is depending on the ability of adversary to get the location of the BS. These configurations are used to show the effectiveness of proposed scheme and the effect of cell size on anonymity metrics.

V. RESULTS

The experimental results is compared with and without the implementation of BBS and also with BAR approach proposed in [12]. The figure 1 & 2 show the entropy measurement for 20 cells with an interval of 1 hour. The total time period for this experiment is 10 hours. The minimum degree of anonymity is 0.6. The result shows that entropy of the system is increase when the BBS and Bar approach is applied because the traffic is more spread due to the retransmission nature of the BS. The average increase in the entropy is 23% when the BBS is employ while the average increase in the entropy is 19% when the BAR approach is applied. Although the BAR approaches increase the anonymity of the BS but it is less efficient then BBS.

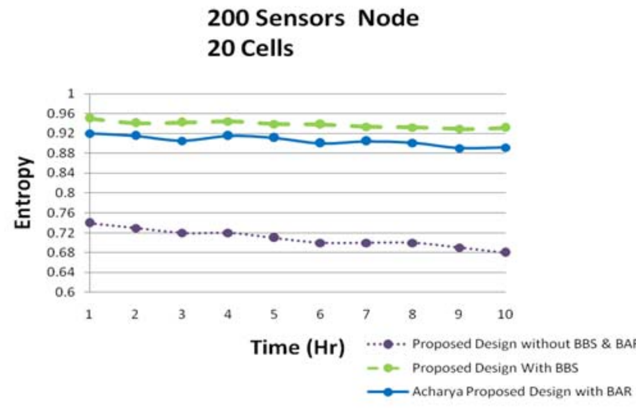


Figure 1 Entropy of 200 Sensor Nodes for 20 Cells

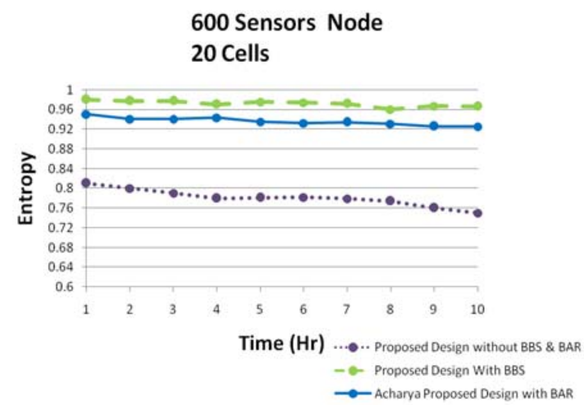


Figure 2 Entropy of 600 Sensor Nodes for 20 Cells

The same experiment is taken on 40 & 60 cells respectively, which verify that the BBS approach is efficient and increases the BS anonymity as compare to BAR approach as shown in the figures 3 to 6. The results show that the size of cell does not affect the entropy of the system. The entropy of the system is only affected by the number of sensors nodes. If the numbers of sensor node increases the entropy of the system also increases, the traffic is distributed on many path which complex the traffic analysis.

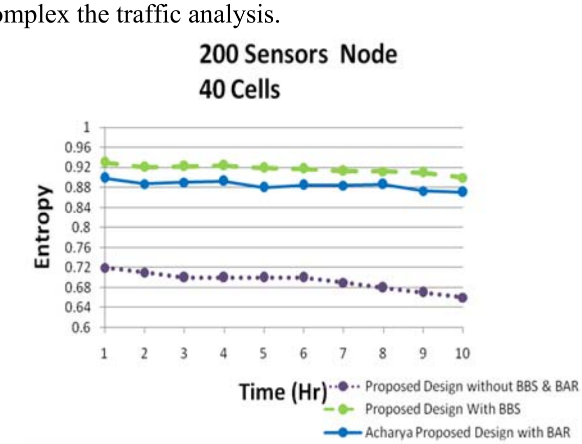


Figure 3 Entropy of 200 Sensor Nodes for 40 Cells

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

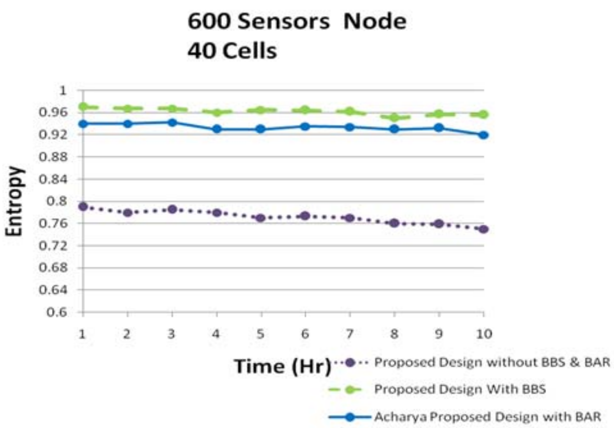


Figure 4 Entropy of 600 Sensor Nodes for 40 Cells

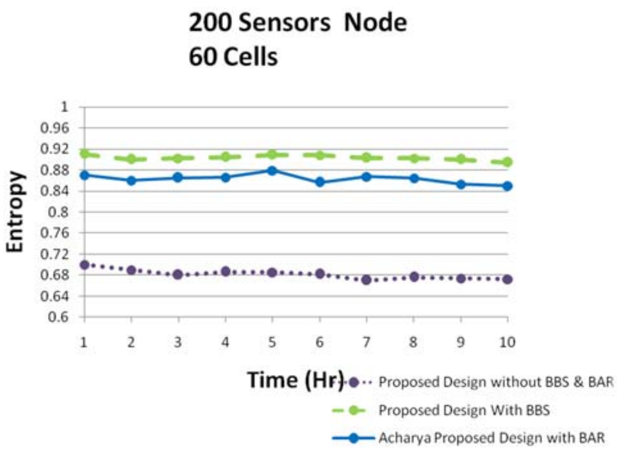


Figure 5 Entropy of 200 Sensor Nodes for 60 Cells

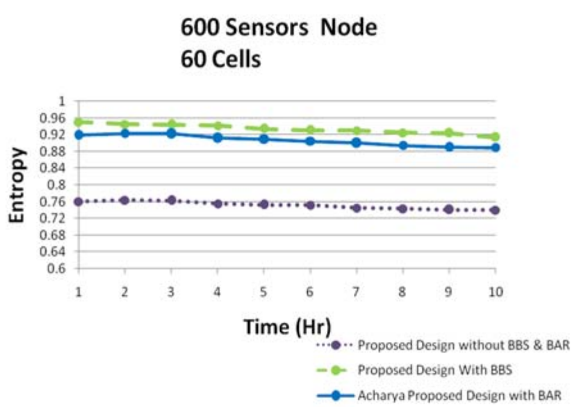


Figure 6 Entropy of 600 Sensor Nodes for 60 Cells

The figures 1 to 6 also show that the entropy of the system is decreases in all three scenarios over a period of time. The adversary takes enough information and will be closer to the BS. The cell which contains the BS is predictable for the adversary because the adversary is more familiar with traffic of the network.

The GSAT measurement is employed on 20, 40 and 60 cell respectively. The GSAT assign the probability to the cell which contains the BS based on the number of time an adversary visit to the cell. If the adversary consecutively visits to a cell, the probability of that cell increases. The adversary observes the traffic from 10 to 15 minutes in their surroundings at the start of simulation and move towards the cells containing the higher transmission. Experiment result show that average decrease in the probability is 11 % when the BBS is applied while the average decrease in the probability is 7 % when the BAR approach is applied. Figure 7 & 8 confirm the efficiency of BBS approach.

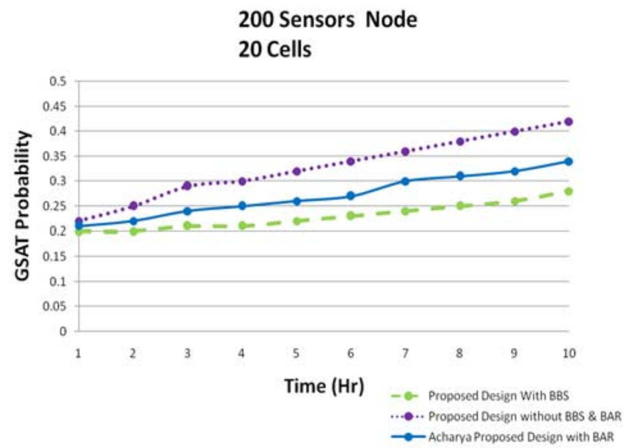


Figure 7 GSAT Probability of 200 Sensor Nodes for 20 Cells

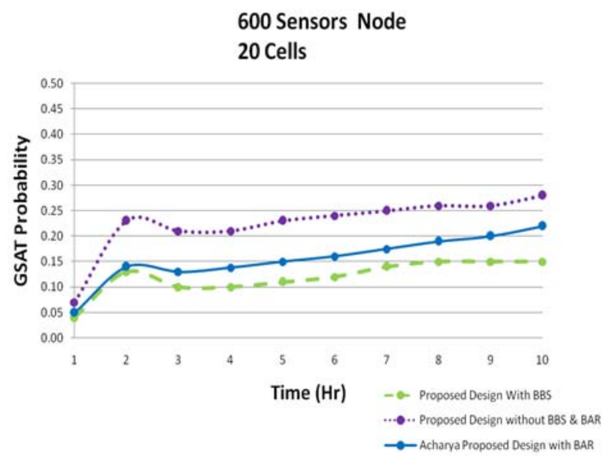


Figure 8 GSAT Probability of 600 Sensor Nodes for 20 Cells

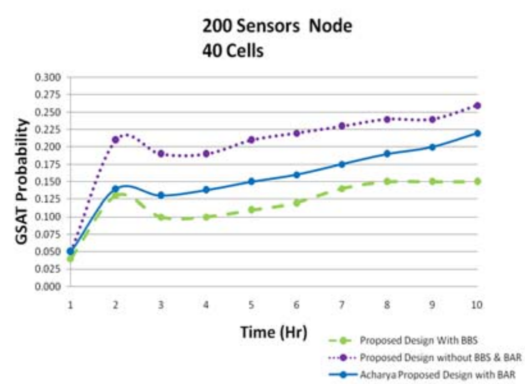


Figure 9 GSAT Probability of 200 Sensor Nodes for 40 Cells

The probability assign to the BS cell is dependent on the number of cell an area is divided. The area is divided into small number of cells than the probability assign to the BS cell

is higher. If the area is divided into large number of cell than the probability assign to the BS is lower. The figures also show that the probability of GSAT matrices is increase over period of time because the adversary takes much more knowledge to predict the traffic pattern.

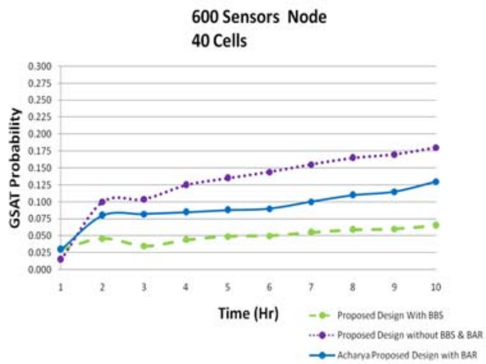


Figure 10 GSAT Probability of 600 Sensor Nodes for 40 Cells

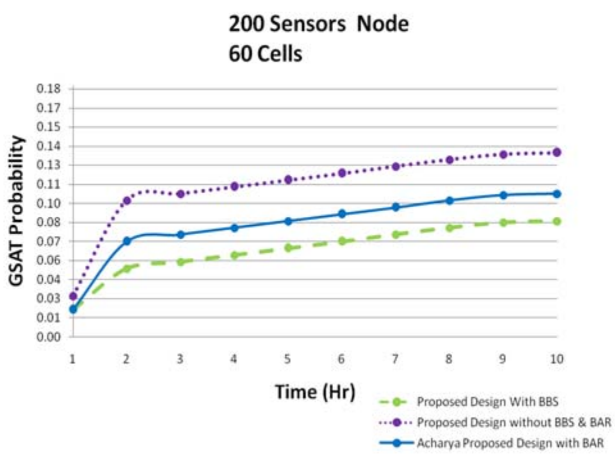


Figure 11 GSAT Probability of 200 Sensor Nodes for 60 Cells

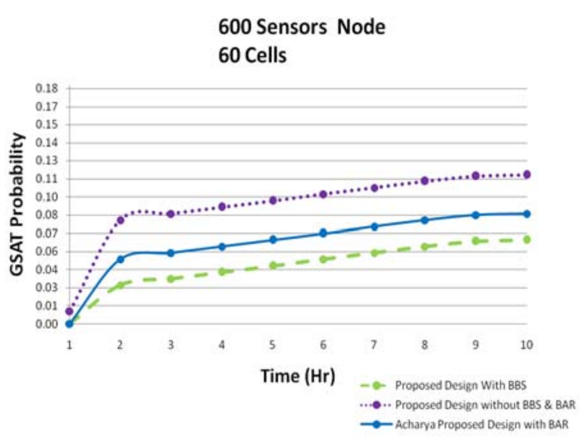


Figure 12 GSAT Probability of 600 Sensor Nodes for 60 Cells

The Results show that both approaches boost the anonymity of the BS and also confirm that after a period of time the anonymity of the BS is decrease. The adversary gets the efficient knowledge to destroy the BS because adversary continuously monitors the traffic pattern. Once the adversary destroy the BS the whole network is useless. BAR approach only enhance the anonymity of BS but not provide a backup facility when BS is destroyed. BBA enhanced the anonymity of the BS more efficiently than BAR approach and provide backup facility when the BS is destroyed.

Both base stations will have the ability to behave like sensor nodes by transmitting signals. BS will transmit "interest" to sensor nodes in direct coverage area for collection of data from environment. The sensor nodes will collect data from environment and transmit it to the BS. BS will process the data and at the same time will send it towards BBS after regular intervals. BBS receive data from BS and from other sensor nodes. The data collected from other sensor nodes by BBS is sent to BS. BBS will also send a message to BS on regular intervals and in response to the message BS will send data towards BBS.

The two way communication will symbolize the live status of both base stations. Since there is no end point in the network therefore adversary will not be able to locate base station. Even if adversary finds the location of base station and destroy it or if the base station is destroyed by energy failure [battery power failure], the BBS will take charge and will continue the operation of network smoothly. Resultantly the downtime of network will also be minimized or negligible.

VI. CONCULUSION

A method has been proposed and demonstrated employing two BS in a hidden location with the aim of promoting anonymity. One of the BS is working as a BBS while the other is as normal BS. Both behave as sensor node to complex data traffic analysis for the adversary. The BBS not only send the collected data from the environment to the BS but also receives the data from the BS to complex data traffic. An adversary use entropy and GSAT matrices to find the location of the BS. Due to confusion the adversary requires more efforts to expose the BS location. The increase in graph proves the efficiency of the proposed approach. This technique boosts the BS anonymity and provides backup facility. The adversary is still capable of detecting BS location after a period of time by utilizing its resources. However if the adversary finds the location successfully and destroys the BS then the BBS takes over the charge and runs the network smoothly.

In future, it would be interesting to enhance the proposed scheme by using the BBS as a mobile node. The BBS will change its position regularly which would further challenge the traffic analysis.

VII. REFERENCE

[1] T. Bokareva, W. Hu, S. Kanhere, B. Ristic, and N. S. Wales, "Wireless Sensor Networks for Battlefield Surveillance," *Signal Processing*, no. October, pp. 1–5, 2006.

[2] M. Hefeeda and M. Bagheri, "Forest Fire Modeling and Early Detection using Wireless Sensor Networks,"

Ad Hoc Sens. Wirel. Networks, vol. 7, no. August 2003, pp. 169–224, 2009.

[3] B. J. Ko, C. Lu, M. B. Srivastava, J. a Stankovic, F. Ieee, A. Terzis, and M. Welsh, “Wireless Sensor Networks for Healthcare,” vol. 98, no. 11, 2010.

[4] M. Winkler, K. Tuchs, K. Hughes, and G. Barclay, “Theoretical and practical aspects of military wireless sensor networks,” *J. Telecommun. Inf. Technol.*, vol. 2, pp. 37–45, 2008.

[5] J. Deng, R. Han, and S. Mishra, “Decorrelating wireless sensor network traffic to inhibit traffic analysis attacks,” *Pervasive Mob. Comput.*, vol. 2, no. 2, pp. 159–186, 2006.

[6] X. Luo, X. Ji, and M.-S. Park, “Location Privacy against Traffic Analysis Attacks in Wireless Sensor Networks,” *2010 Int. Conf. Inf. Sci. Appl.*, pp. 1–6, 2010.

[7] B. Di Ying, D. Makrakis, and H. T. Mouftah, “Anti-traffic analysis attack for location privacy in WSNs,” vol. 2014, no. 1, pp. 1–15, 2014.

[8] J. Xu, N. Jin, X. Lou, T. Peng, Q. Zhou, and Y. Chen, “Improvement of LEACH protocol for WSN,” no. Fskd, pp. 2174–2177, 2012.

[9] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “Wireless sensor networks: a survey,” *Comput. Networks*, vol. 38, no. 4, pp. 393–422, 2002.

[10] S. Mahmud, H. Wu, and J. Xue, “Efficient energy balancing aware multiple base station deployment for WSNs,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6567 LNCS, pp. 179–194, 2011.

[11] S. J. Anandh, “Energy Efficient Routing Strategies for Clustered Wireless Sensor Networks : An Analytical Framework,” vol. 74, no. 8, pp. 19–27, 2013.

[12] U. Acharya and M. Younis, “Increasing base-station anonymity in wireless sensor networks,” *Ad Hoc Networks*, vol. 8, no. 8, pp. 791–809, 2010.

[13] A.-S. Abuzneid, T. Sobh, M. Faezipour, A. Mahmood, and J. James, “Fortified Anonymous Communication Protocol for Location Privacy in WSN: A Modular Approach,” *Sensors*, vol. 15, no. 3, pp. 5820–5864, 2015.

[14] J. Chen, X. Du, and B. Fang, “An efficient anonymous communication protocol for wireless sensor networks,” 2009.

[15] B. Alomair, A. Clark, J. Cuellar, and R. Poovendran, “Statistical Framework for Source Anonymity in Sensor Networks Network Security Lab (NSL), Electrical Engineering Department , Siemens Corporate Technology , M ¨.

[16] Z. Ren and M. Younis, “Exploiting architectural techniques for boosting base-station anonymity in wireless sensor networks,” *Int. J. Sens. Networks*, vol. 11, no. 4, p. 215, 2012.

[17] N. S. Jyothi, “Location Privacy Methods in Wireless Sensor,” vol. 3, no. April, pp. 160–165, 2013.

[18] R. El-Badry and M. Younis, “Providing location anonymity in a multi-base station wireless sensor network,” *2012 IEEE Int. Conf. Commun.*, no. July, pp. 157–161, 2012.

[19] E. C.-H. Ngai, “On providing sink anonymity for sensor networks,” pp. 269–273, 2009.

[20] Xinfeng Li. Xiaoyuan Wang, Nan Zheng, Zhiguo Wan, and Ming Gu, “Enhanced Location Privacy Protection of Base Station in Wireless Sensor Networks,” *Mob. Ad-hoc Sens. Networks, 2009. MSN ’09. 5th Int. Conf.*, 2009.

[21] J. R. Ward and M. Younis, “Examining the Effect of Wireless Sensor Network Synchronization on Base Station Anonymity,” *2014 IEEE Mil. Commun. Conf.*, pp. 204–209, 2014.

[22] W. Conner, T. Abdelzaher, and K. Nahrstedt, “Using Data Aggregation to Prevent Traffic Analysis in Wireless Sensor Networks,” *Lect. Notes Comput. Sci.*, vol. 4026, p. 202, 2006.

[23] Y. Ebrahimi and M. Younis, “Increasing Transmissio on Power for Higher Base-sta ation Anonymity in W Wireless Sensor Network,” *Commun. Soc.*, pp. 2–6, 2011.

[24] M. Shao, Y. Yang, S. Zhu, and G. Cao, “Towards statistically strong source anonymity for sensor networks,” *Proc. - IEEE INFOCOM*, pp. 466–474, 2008.

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 6, June 2016

Neural Feed Forward Fault Tolerant backbone tree construction to increase the lifetime of wireless Sensor Network

K Vimal Kumar Stephen[1], Mathivanan V[2]

1. Research Scholar, Computer Science and Engineering Department
AMET University, Chennai, India,

2. Research Supervisor, Computer Science and Engineering Department
AMET University, Chennai, India.

ABSTRACT - **In the recent times, the demands of Wireless Sensor Networks (WSN) increase the challenges in terms of scalability and energy efficiency. One of the key challenges in the wireless sensor network is how to prolong the lifetime of the network. To improve the lifetime of the sensor, static and movable mobile sinks are deployed. Movable sinks are used to receive sensed data from the sensor where it is located. The static mobile sinks act as a trusted third party for computing and distributing keys between sensor nodes and the clusters. It is not necessary to chose new cluster head often because of trusted third party sink, performs all the computations of cluster head. The energy is retained when computation is reduced in cluster head thereby increases the life time of the particular cluster. Feed forward Back propagation algorithm is proposed using adaptive learning in neural networks followed by link aware routing. This algorithm deals with fault tolerant backbone tree construction for data transmission whereas it produces optimal path for the sink to transmit data. Since the optimal path is established, the life of the sink also to be prolonged thereby increase the overall network lifetime. Result shows that the lifetime of the network is improved and energy depletion is reduced.**

Keywords – Sensor Networks, mobile sink, clusters

I. INTRODUCTION

Network lifetime has become an important challenge for evaluating sensor network [1, 2, and 3]. Sensor coverage, connectivity and node coverage play a key role in deciding the lifetime of the sensor network. There are also several other factors that determine the lifetime of a sensor network like mobility, heterogeneity, quality of service and completeness. Many routing algorithms were proposed for energy efficiency to improve the lifetime of wireless sensor network (WSN). In the scenario of energy efficiency, wireless sensor network encounters loss of battery power during communication. Sensor node (SN) senses the data in the environment and transmits to the base station (BS) through the cluster head. Battery is drained when the data is sensed and also during transmission of sensed data. The battery is drained in cluster head during computation of keys and data transmission.

The issue present here is during data transmission from one sensor node to another sensor node, it takes more hops to reach cluster head/other sensor node/base station hence the energy is drained to the maximum. In order to retain the energy of the cluster head, energy efficient cluster based scalable key management technique has been proposed with mobile sinks to increase the lifetime of the cluster head which in turn increasing the lifetime of the network. Also routing path may differ when transmission of data takes place from sensor to BS. If the path is so long and large number of SN involved in transmitting the data and then this process repeats, it leads to energy depletion. The lifetime of the network increases when the route to transfer the data is optimal. The current scenario generally noticed problems in Wireless Sensor Networks are, 1) Security breach needs to be solved, 2) Scalability of sensor node 3) Optimized path for routing and 4) Energy efficiency is still need to be improved. This research helps to overcome the above mentioned problems and issues thereby increases network lifetime and it is energy efficient.

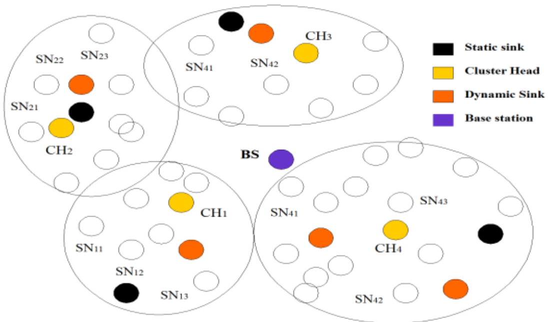


Figure 1 Proposed architecture

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

II. LITERATURE SURVEY

Many efforts have been made for energy efficiency of WSNs [1, 6] and Hop selection strategies such as one-hop neighboring nodes or multiple links for transmitting data have been focused by many researchers [7–9]. Some researchers focused on scheduling the nodes. An efficient routing metric used in taking the selection of next hop in routing. Recently, two categories of routing methodologies were attractive, Cluster based and Virtual backbone based. Cluster based protocols [4, 5, 6, 7] uses single-hop communication model, each sensor node sends packet to its cluster head directly in a single hop and the cluster head transfers the sensed data to the cesspool. The existing solution for clustering requires maintenance to reorganize the clusters due to mobility and node failures.

Virtual backbone [8, 9, 10] to organize node in a better way can be produced by different algorithms. A backbone is a subset of active nodes is able to perform a special task and serve nodes which are not in the backbone. A backbone reduces the operating cost involved in the communication between sink and other sensor nodes, decreases the overall energy consumption of each parcel and also increases the network lifetime in WSN.

A new algorithm called Connecting Dominating Set Augmentation (CDSA) [11] is proposed to protect the failure of one wireless node by constructing a 2-connected virtual backbone. The size of the CDSA constructed 2-connected backbone has guaranteed the quality within a constant factor of the optimal 2- connected virtual backbone size. CDSA can build a 2-connected virtual backbone with only small overhead through simulations.

The problem of constructing fault-tolerant CDS [12] in homogeneous wireless networks is investigated, which is abstracted as the minimum connected dominating set problems. A constant factor polynomial-time approximation algorithm is computed and this algorithm works for any abstract graph without the information of geometric coordinates of the input graphs. The property of UDG is used in the analysis part to get a constant approximation. Also investigated a constant factor approximation algorithm for k>=3 and m>=1 in a disk graph. Multipath transmission is proposed [13] enables fast transmission and coverage metric proved in this research yields maximum lifetime in the network**.**

III. NEURAL FEED FORWARD FAULT TOLERANT BACKBONE TREE CONSTRUCTION

Initially, sensor nodes are dynamic after deployment during the network operation. Cluster head (CH) is also dynamic which can be chosen by cluster head formation algorithm. Initially the cluster head is chosen among on sensor node which is having highest battery power. The new cluster head is selected by the algorithm only when the already exiting cluster head reaches the threshold value.

There are scenarios when energy depletion takes place. They are 1) When any of the sensor node joins or leaves the network group, it is necessary to care about secrecy. During each joins and leaves, enough computation needs to be done for authentication, encryption of data, secure transmission, etc [1]. Each time computation such as identity generation secure key generation is being done by Cluster Head drains the energy present in the short span of time. 2) If the SN is far away from the base station, CH needs to transmit the data through multiple nodes. If it happens again and again hence there is a more chance of energy drain.

To avoid this, two or more movable sinks and one static sink is placed in the network. If any of the SNs is ready to transmit the data, the movable sink is readily available moving around the SNs in the clusters gets the data and transmits to the BS. Hence loss of energy is reasonably avoided in the CH because sink gets the data directly from the SNs, move towards BS and delivers the data. This sink helps in saving battery power of CHs and all other SNs.

The static sink is a trusted node deployed for computing the above mentioned key management system so that the computation overhead is reduced drastically in the CH. It acts as a proxy for CH in such a way that it prevents energy loss. This static sink does the computations upon the request of CH about the join and leaves of SNs. This static sink concept helps to prolong the life of CH to the extent. The cluster head election algorithm is executed once the CH reaches the energy level below threshold level, new CH is elected based on highest energy level among the SNs.

The potential problem in the previous defined protocols is that once the optimal route is determined, sending data through the same path leads to energy depletion of all the nodes in the path may lead to network partition. It is done to have Multipath Data Transmission which consumes less energy and maximum coverage. These paths are chosen based on the probability on how low the energy consumption of each path is. The optimal paths are chosen accordingly due to the probabilistic choice of routes since it continuously evaluates along different routes.

Let Ni , Nj , Ns , ND are the intermediate, source and destination nodes, while R_C is the routing metric. R_C is the value of routing cost initially set to zero but it is updated when the data

transmission takes place along the optimal path. To maintain the coverage and connectivity, every intermediate node forwards all the requests only to the neighbor nodes that are closer to the N_S than itself and farther away from the N_D. Routing cost of each path from source to destination and intermediate nodes is updated in the sink for further optimal path finding. The path having the high cost is discarded according to the defined metric. Each path is assigned the probability for successful transmission according to routing cost metric given below

$$R\_C = \left[\frac{E_i^D}{E_t(S_i, S_j) + E_r(S_i, S_j)} * \frac{1}{R} \right]$$

This multi path transmission is done with the help of Fault tolerant Virtual Backbone Tree (FTBT) to reduce the energy consumption for a packet, thereby increases network lifetime, doesn't drain any particular node quickly and also maintains N-of-N lifetime.

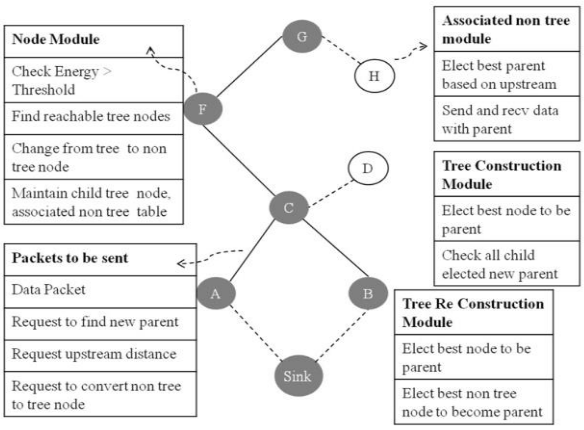


Fig.2. Feed Forward Backproagation Virtual Tree System

Virtual backbones provide an infrastructure for communicating efficiently to the sink node [14]. Nodes are elected as tree nodes and non tree nodes. All communication between particular sensor nodes to the sink node happens through the tree nodes. There are various strategies to elect the tree nodes. In EVBT [15], sink node transmits BCR (Broadcast Request Packet) packet to all nodes in its sensing range. The nodes on receiving this packet compute its fitness factor and time delay (t_d) are inversely proportional to each other. The node waits until t_d is expired. If the node has received one more BCR request in this interval, the node becomes non tree node and elects the nearest tree node as its upstream link. Else the node becomes a tree node and a part of virtual backbone. Electing the nearest tree node for the upstream link does not involve much computation, but it is not the shortest route to the sink. Here, Sink node and sensor nodes that are classified as tree and non-tree nodes.

A sensor has comparatively high energy and it performs sensing, sending and receiving data called tree node. All data from non-tree node to sink need to be transferred with minimum energy to prolong the lifetime of the sensor. Since each node periodically checks its energy, a tree node becomes a non-tree node when a node's energy reaches its threshold level. Two Packets to be sent includes data packets and request packet. Data packet holds data that is sensed by the sensor nodes, which is sent to the sink via the virtual backbone tree. Request packet is sent for finding new parent and upstream distance. How tree and non tree node elected depends on the lifecycle of the node.

It is necessary to eliminate the sensor which is having minimum energy below threshold value and retain the link between the nodes by constructing the virtual backbone tree. It is done using neural network back propagation feed forward algorithm. It retains the network structure and its link. The aim is to construct NNs with nearly minimum number of hidden neurons. In the neural network, the input layer is fully connected to the hidden layer or neurons which is also fully connected to the output layer. The output o_j of the j_{th} neuron is given by

$$O_j = f\left(\sum_{i=0}^{k} w_{ij} O_i\right)$$

where w_{ij} is the synaptic weight with respect to the connection from the i^{th} neuron in the previous layer to the j^{th} neuron, K is the number of neurons that feeds the j^{th} neuron, O_i is the output of the i^{th} neuron, **f** is the sigmoid activation function given by **f(x)** = $1/(1 + e^{-x})$. The hidden neurons can be added when needed one by one, each receives a connection from each of the network's inputs. All the input-to-hidden and hidden-to-output weights are trained repeatedly, not only the hidden-to-output weights.

All the sinks initially trained with neural network back propagation feed forward algorithm which minimizes the link break probability in order to find its new optimal route to transmit the data instead of sending in the same route. After training, it is generalized to train a pool of candidate neurons to select the best neuron among the pool. Each candidate with different set of initial weight is temporarily connected to the output of every input neuron. Its output is also temporarily connected to every neuron in a virtual output layer, where **a** virtual output layer is a temporal layer of the same size **as** the original output layer. Hence, each time a sink constructs path from the sensors where it is receiving data, if the path is not optimal and if there is maximum of non tree nodes then it back propagates thereby finds optimal path to destination.

IV. BACKBONE CONSTRUCTION

In the sensor network, energy level of each node varies. Initially, all the nodes with an energy level

greater than the threshold (T) are temporarily marked as tree nodes. A tree is constructed by the sink connecting all these nodes. If a tree node has too many dependents or reached its threshold value, it will become a hotspot and will lose energy quickly by transferring many packets through it. So one should concentrate on how a node comes to know whether it has many dependents. The approach is to calculate the average number of dependents for all other reachable tree nodes which are in its sensing range.

If the number of dependents of the present node is two times the average dependents, it must try to reduce its number of dependents. It will ask its dependent child tree node and associate non tree node to check if it can find a new parent tree node. This node remains parent for only those nodes which do not find a new parent. Though the backbone tree needs to be constructed from a minimum number of nodes covering the entire network, having more dependents for a particular tree node will have severe impact on its energy. Tree node whose energy is approaching threshold should reduce its number of dependents appreciably and possibly try to become a non tree node.

Algorithm 1:

N_i If N_i.Energy> T

$N_i \leftarrow$ TN // temporarily. Form a tree, root: sink and initiated by it

$\forall N_i$

RTN[ ] $\leftarrow$ { N_j==TN ∩ dist(i,j)<S_i }

// let n be the number of reachable tree nodes

Sum $\leftarrow$ 0

$\forall$ RTN

Sum+=RTN→No.of.dependents

Avg=Sum/n

if RTN → No. of.dependents ≥ 2 x Avg || N_i.Energy → T + ε **then** find a suitable parent

A) Finding a suitable parent for child tree nodes

One can finds all its reachable tree nodes of this child tree node which lie within its sensing range for all child tree nodes. If there is only one node in its range, then it is chosen as its parent. The node with highest fitness factor is chosen as its parent if there is more than one tree node in its sensing range. If the parent is undefined, the child tree node waits so that any of its reachable sibling tree nodes of its earlier parent has chosen a new parent. Then the sibling is chosen as its parent. If it couldn't find a parent, then node N_i remains as a tree node, so that the network sustains.

B) Finding parent for associated non tree nodes

The reachable tree nodes which lie in its sensing range for each non tree child node are found. If the number of such nodes is just one, then it is chosen as its parent directly else compare the upstream distance for each tree node, assuming that as its parent, now node with least upstream distance is chosen as its new parent node. Node N_i must check if its entire child nodes (tree and non-tree) are assigned a new parent, then N_i becomes a non-tree node else the node remains as tree node. In short the parent for child tree nodes are chosen based on maximum fitness value, and the parent for child non-tree node is chosen based on minimum upstream distance.

V. BACKBONE RECONSTRUCTION

Reconstruction of a tree is needed when node fails due to hardware error or complete drain of energy. Tree nodes periodically check whether its energy falls below threshold energy. If so the node becomes a non-tree node. Let T be a tree node which fails, all its child tree nodes are assigned to a new parent. A child tree node finds all the tree nodes in its sensing range. If there is only one node, it is made as its parent, else a parent is chosen based on the highest fitness factor and its other parameters like upstream distance and angle. If there is no other tree node in its sensing range, it checks all non tree nodes, which are in its sensing range and selects the one with best fitness factor and minimum upstream distance and converts it from non tree to tree node. The pre-condition is the selected non tree must have energy greater than threshold. If a non-tree fails, then there is no breakage in the tree structure, hence that node is removed from the tree formed and considered to be dead.

VI. RESULT AND DISCUSSION

Table 1. Simulation Parameters

Parameters	Value
Region in radius	400 m*400 m
Sensing range of nodes	60 m
No. of Nodes	200
Initial energy per node	5 J
Network bandwidth	2 Mbps/s
Power to run the transmitter/receiver circuitry	70 nJ/bit
Power for the transmit amplifier to achieve an acceptable SNR (Signal to Noise Ratio)	120 pJ/bit/m2
Data packet Size	4096 bits
Control packet size	20 bits
Data transmission rate	4096 bits

The virtual backbone tree construction is proposed using neural network, an adaptive learning. Weights of neurons are assigned according to the residual energy of the nodes in the network. A coverage aware routing metric is

included in this scheme to choose the best route. Since it is multipath transmission, one of the routes is chosen from the routes found and decided. The data transmission is performed using the defined metric. The results obtained in the proposed scheme are quite effective and it delivers more than 95% of the packets to their destination with increased network coverage. Although with an increase in network coverage, the number of alive nodes decreases with respect to coverage and connectivity.

Table 2: Transmission range, Average number of nodes vs Average number of dependents for each tree node

Transmission range in meter(m)	Average number of nodes	Average number of dependents for each tree node
20	0	3.6
25	20	5.66
30	40	7.94
35	60	10.44
40	80	13.06
45	100	16.49
50	120	18.52
60	140	21.09
70	160	24.14

VII. CONCLUSION

This paper proposes a fault tolerant Feed Forward back propagation network algorithm which emphasizes the lifetime maximization. This virtual backbone tree is flexible and duration of the network for a longer period of time is maintained, hence, N – N lifetime is achieved through virtually connected sink. Multipath transmission is enabled to improve the performance of the network and fast data transmission. Deployment of static and dynamic sink in the network helps to prevent sensor nodes and cluster head form energy drain in turn it increases the lifetime of the sensor network. This leads to negligible storage overhead and communication overhead thus it saves energy. Results proved that the proposed method gives better performance and achieved the major challenges in wireless sensor networks.

References

[1] Jae-Hwan Chang and Leandros Tassiulas," Maximum Lifetime Routing in Wireless Sensor Networks",IEEE/ ACM Transactions on Networking, VOL. 12, NO. 4, AUGUST 2004,DOI:10.1109/TNET.2004.833122

[2] Md Nafees Rahman, M A Matin,Md Nafees Rahman, "Efficient Algorithm for Prolonging Network Lifetime of Wireless Sensor Networks", Tsinghua Science and Technology , December 2011, Volume 16, Number 6,pp561-568

[3] YoungSang Yun, Ye Xia, Behnam Behdani, and J. Cole Smith," Distributed Algorithm for Lifetime Maximization in a Delay-Tolerant Wireless Sensor Network with a Mobile Sink", IEEE Transactions on Mobile Computing, VOL. 12, NO. 10, OCTOBER 2013, DOI: no. 10.1109/TMC.2012.152.

[4] A. Manjeshwar and D. P. Agarwal, "TEEN: a Routing Protocol for Enhanced Efficiency in Wireless Sensor Networks,"1st International Workshop on Parallel and Distrib. Computing, 2001, DOI: 10.1109/IPDPS.2001.925197.

[5] A. Manjeshwar and D. P. Agarwal , "APTEEN: A Hybrid Protocol for Efficient Routing and Comprehensive Information Retrieval in Wireless Sensor Networks," Parallel and Distributed Process. Proc. Symp, pp. 195–202, 2002 ,DOI: 10.1109/IPDPS.2002.1016600.

[6] Deng.S , "Mobility-based clustering protocol for wireless sensor networks with mobile nodes," IET Wireless Sensor Systems, Vol.1, Iss. 1, pp. 39-47, 2011.

[7] J.Akbar Torkestain,M.R.Meybodi, "An intelligent backbone formation algorithm for wireless as networks based on distributed learning automate,"Computer Networks,Vol 54, pp.826-843,2010, DOI: 10.1016/j.comnet.2009.10.007.

[8]K.M.Alzoubi,"New Distributed algorithm Connected Dominating Set in Wireless Ad Hoc Networks," Proceedings of the 35th Annual Hawaii International conference on System Sciences(HICSS'02), Vol.9.pp.297,2002.

[9] R. Xie,D.Qil , " A novel distributed MCDS approximation algorithm for wireless sensor networks,"Mobile & Wireless Communication,Vol.9,Issue 3, pp.427-437,2009.

[10] I.FAkyildiz , "A Survey on Sensor Networks," IEEE Communication Magazine, vol. 40, pp. 102-114,2002 DOI: 10.1002/wcm.547.

[11] Feng Wang, Member, IEEE, My T. Thai, Member, IEEE, and Ding-Zhu Du, "On the Construction of 2-Connected Virtual Backbone in Wireless Networks", IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, VOL. 8, NO. 3, MARCH 2009.

[12] Wei Wang, Donghyun Kim, Min Kyung An, Wei Gao, Xianyue Li, Zhao Zhang, and Weili Wu," On Construction of Quality Fault-Tolerant Virtual Backbone in Wireless Networks", IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 21, NO. 5, OCTOBER 2013.

[13] Neeraj Kumar, Manoj Kumar and R.B. Patel, "Coverage and Connectivity Aware Neural Network Based Energy Efficient Routing in Wireless Sensor Networks", Journal on Applications of Graph Theory in Wireless Ad hoc Networks and Sensor Networks, Vol.2,No.1, March 2010.

[14] Bosheng Zhou,"An energy-aware virtual backbone tree for wireless sensor networks,"IEEE Global Telecommunications Conference, pp. 1212-1215,2005 DOI: 10.1109/GLOCOM. 2005.1577373.

[15] Kim, Jaekwang and Lee Jee-Hyongng, "ViTAMin: A Virtual Backbone Tree Algorithm for Minimal energy consumption in wireless sensor network routing," International Conference on Information Networking (ICOIN), pp.144, 149,2012 DOI: 10.1109/ICOIN.2012.6164366.

A Efficient Neural Network Model for Software Effort Estimation

[1] Nesa Khandoozi GholiAbad, [2] Sanaz Khandoozi GholiAbad
[1] Islamic Azad University, Gorgan branch, Iran.
[2] Islamic Azad University, Sari branch, Iran.

Abstract— **Software development effort estimation is the process of predicting the effort required to develop a software system. Estimating development effort accurately in the early stage of software life cycle plays a crucial role in effective project management. Effort estimation is a key factor for software project success, defined as delivering software of agreed quality and functionality within schedule and budget. Traditionally effort estimation has been used for planning and tracking project resources. It has become an important task. This paper proposed a neural network model for software effort estimation. This model has 3 layers. The train, validation and test data used are from COCOMO data set. Inputs and targets data randomly divided in train (60 %), validation (20%) and test (20%) group. When the number of neurons in hidden layer was 20, Number of training samples was 37, number of validation samples was 13 and number of testing samples was 13, the network has best performance. In this case, the value of training, validation and testing MSE was 0.01044, 0.0475 and 0.0375 respectively and value of training, validation and testing R was 0.9167, 0.7741 and 0.7410 respectively.**

Keywords- Software Engineering, Effort Estimation, Artificial Neural Network

I. INTRODUCTION

Software development effort estimation is the process of predicting the most realistic amount of effort required to develop or maintain a software project based on information collected in the early stage of a software project. Software development effort estimation is the process of predicting the most realistic amount of effort (expressed in terms of person-hours or money) required to develop or maintain software based on incomplete, uncertain and noisy input. Effort estimates may be used as input to project plans, iteration plans, budgets, and investment analyses, pricing processes and bidding rounds. Those responsible for effort estimations are usually the project managers [1]. Depending on the chosen effort estimation method, they can estimate alone or with expert advice from developers, designers and testers. Other people that need most the effort estimation are project owners and sales. Most of the times, your effort estimation may be challenged by sales or management teams. Sales people want low cost. This means low effort estimation; you want more resources and your most valuable resource might be time. Also, you know that everyone will be happy if you finish earlier and none if you finish later. In addition, developers and designers when giving estimates have in their minds the possibility to be pressed to finish tasks in strict deadline… and, for sure, they don't want the pressure, so, most commonly, they will take the worst case when estimating [2].

Software effort estimation has been an important issue for almost everyone in software industry at some point. Effort estimation is essential for many people and different departments in an organization. Also, it is needed at various points of a project lifecycle. Presales teams need effort estimation in order to cost price custom software and project managers need it in order to allocate resources and times plan a project. Usually, software development is priced based on the person days, it requires in order to be built, multiplied by a daily person day rate. Without effort estimation pricing is impossible. Also, in order to plan a project and inform the project owners about deadlines and milestones you have to know how much effort the job requires. Finally, initial effort estimation shows if you have the resources to finish the project within customer or project owner predefined time limits, based on your available man power [3].

Most of the research has focused on the construction of formal software effort estimation models. The early models were typically based on regression analysis or mathematically derived from theories from other domains. Since then a high number of model building approaches have been evaluated, such as approaches founded on case-based reasoning, classification and regression trees, simulation, neural networks, Bayesian statistics, lexical analysis of requirement specifications, genetic programming, linear programming, economic production models, soft computing, fuzzy logic modeling, statistical bootstrapping, and combinations of two or more of these models [4]. The perhaps most common estimation methods today are the parametric estimation models COCOMO, SEER-SEM and SLIM. They have their basis in estimation research conducted in the 1970s and 1980s and are since then updated with new calibration data, with the last major release being COCOMO II in the year 2000.

Formal estimation approaches usually exploit training data about past projects to build an estimation model which is then used to predict the effort for a new project model takes as input a set of predictors and returns a scalar value that represents the effort estimated to develop a new software system having the characteristics captured by the predictors. This model can be described by the following equation: EstimatedEffort = c1 op1 f1 … cn op2n-1 fn op2n C. where fi represents the value of the

ith project feature and ci its coeffcient, C represents a constant, while opi represents the ith mathematical operator of the model [5].
The estimation approaches based on functionality-based size measures, e.g., function points, is also based on research conducted in the 1970s and 1980s, but are re-calibrated with modified size measures and different counting approaches, such as the "use case points" in the 1990s and COSMIC in the 2000s. In this paper we propose a fuzzy neural network method for software effort estimation. This method is based on fuzzy logic and neural network. This paper organized as follow: Section 1: Introduction, section 2: related works, section 3: COCOMO model, section 4: data collection, section 5 proposed method and section 6: conclusion.

II. RELATED WORKS

Several different methods for automating software cost/ effort estimation have been proposed . During the last 30 years, a number of formal models for software effort estimation have been proposed such as Cocomo [6], Cocomo II [7], SLIM [8], and Function Points Analysis [9]. These models have some advantages, providing a formulaic underpinning of software effort estimation [10]. The Cocomo I model takes the following form:

$$Effort = \alpha * size^{\beta} \prod_{i=1}^{15} EM_i \tag{1}$$

where α and β are two factors that can be set depending on the details of the developing company and EMi is a set of effort multipliers such as Acap, pcap, aexp, Modp, tool, vexp, lexp, Sced, stor, data, time , turn, virt, cplx, rely.

In [11] has utilized Genetic Algorithm (GA) for optimized value of the parameters of COCOMO model. One of the problems of COCOMO model is identification of the optimized value for parameters. According to the results of the experiments, it is possible to say that better effort estimation could be gained via GA. Researchers [12] have used Fuzzy Logic (FL) for SEE in software projects. They have introduced SCE as one of the challenges and the important activities in software development. The suggested method of them shows that the use of FL is a model in software development. For the experiments results, 15 projects of the KEMERER projects set were used. According to the results, it is possible to conclude that the Mean Absolute Relative Error (MARE) and PRED (n) (the evaluation factor) is better in the proposed method is better than the algorithm methods. The cost function has many parameters in software projects. Some of the factors of software process which have direct size on cost estimation are Line of Code (LOC) and KLOC. The results of the experiments of them show that the MARE percent is more accurate using the FL. In [14] GA utilized to SEE. The accuracy of the effort estimation adds to the validity of the software projects and the project manager would be able to manage them better. They have showed that using GA, the Magnitude Relative Error (MRE) rate has decreased in comparison to COCOMO model. The results of them on COCOMO dataset show that GA is better in estimation in comparison to COCOMO model. Multi objective particle swarm optimization algorithm have utilized [13] for optimization of COCOMO model parameters to minimize the MARE. For more study on the results, they have tested the suggested model on any project. According to the results of the experiments, the MARE value for small projects in COCOMO model is 16.1306 %, and is 9.0143% in suggested model, and is 18.1548% in large projects in COCOMO model and is 20.9717% in proposed model. The results of the experiments show that the suggested model is more efficient. Researchers [14] have SCE using the soft calculations techniques. They have utilized FL and PSO algorithm combination for cost estimation. They have used 30 projects of NASA software project dataset for the results of their experiments. According to the results of the paper, the proposed model could have estimated better in comparison to the various models and make Mean Magnitude of Relative Error (MMRE), 7.512%. In [15], it has evaluated and tested GA using SEE. In this research, the COCOMO model is less accurate in comparison to the artificial intelligence models in SEE. So, it is tried make the parameters in the proposed model more optimized and also make the effort estimation more accurate. In this reference the NASA software project dataset is used for the results of the experiments. According to the results, the suggested model could be better in estimation and reduce the MMRE to 0.2298% in comparison to various models.

A study [20] using a total of thirteen data sets and eight approaches including MLPs, RBFs, RTs and EBA showed that Bagging+RTs were frequently the best ranked approaches in terms of MAE and,

When they were not ranked best, they rarely performed considerably worse than the best approach for a given data set. Another example of ensemble approach was proposed by Kocaguneli et al. (2012) [16-17]. Their method combines several types of so called solo-methods (combinations of single learners and preprocessing techniques) to perform SEE. They reported that the ensemble presents less instability

Than solo-methods when ranked in terms of the total number of wins, losses and wins − losses considering several different performance measures and twenty data sets. They also reported that the ensembles obtained less loss than other methods. As an additional contribution, their extensive study showed that the non-linear approaches CART (a type of RT) and EBA based on log transformed data can outperform other methods such as linear regression based on log transformed data. However, their approach has high

Implementation complexity and is not fully automated. It requires an extensive experimentation procedure using several types of single learners and preprocessing techniques for creating the ensemble. It consists of selecting the "best" solo-methods in terms of losses and stability to compose the ensemble, by manually/visually checking and comparing their stability. The manual/visual checking process is needed because it is necessary not only to determine what solo-methods have the lowest number of losses (that by itself could be automated), but also to check whether these are the same as the ones comparatively more stable and what level of stability should be considered as comparatively superior.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Jenkins, Naumann and Wetherbe [18] conducted a large empirical investigation in the beginning of the 1980s. The study focused on the early stages of system development. It included development aspects, such as user satisfaction, development time, and cost overruns. They interviewed managers from 23 large organizations and collected data on 72 projects. The average project cost was $103,000, and the average duration was 10.5 months. The study included projects that were considered small, medium and large relative to the organizations standards. A majority of the projects developed new software systems (55%), but redesign (33%) and enhancement (11%) of existing software systems were also represented. The survey measured three success factors; user satisfaction, being “on-time” and being “on-budget”.

Bergeron and St-Arnaud [19] performed a study to identify estimation methods, and to what extent they were used. They also investigated how choice of method, and underlying factors and variables, influenced estimation accuracy. In total, 374 Questionnaires were sent to 152 organizations. The companies each received 1-4 copies of the questionnaire. The 89 responses received came from 67 different organizations. All projects included were larger than 150 person-days.

Heemstra and Kusters [20-22] conducted a survey of cost estimation in Dutch organizations. The goal was to provide an overview of the state of the art of estimation and controlling software development costs. They sent out 2659 questionnaires, and got responses from 598 organizations. Estimation methods, original project Estimates and actual effort were analyzed.

III. COCOMO MODEL

The most fundamental calculation in the COCOMO model is the use of the Effort Equation to estimate the number of person-Months required to develop a project. Most of the other OCOMO results, including the estimates for Requirements and Maintenance, are derived from this quantity [6].

A. Source lines of code

The COCOMO calculations are based on your estimates of a project's size in Source Lines of Code (SLOC). SLOC is defined such that:

- Only Source lines that are DELIVERED as part of the product are included -- test drivers and other support software is excluded
- SOURCE lines are created by the project staff -- code created by applications generators is excluded
- One SLOC is one logical line of code
- Declarations are counted as SLOC
- Comments are not counted as SLOC

The original COCOMO 81 model was defined in terms of Delivered Source Instructions, which are very similar to SLOC. The major difference between DSI and SLOC is that a single Source Line of Code may be several physical lines. For example, an "if-then-else" statement would be counted as one SLOC, but might be counted as several DSI [9].

B. The Scale Drivers

In the COCOMO II model, some of the most important factors contributing to a project's duration and cost are the Scale Drivers. You set each Scale Driver to describe your project; these Scale Drivers determine the exponent used in the Effort Equation.

The 5 Scale Drivers are:

- Precedentedness
- Development Flexibility
- Architecture / Risk Resolution
- Team Cohesion
- Process Maturity

Note that the Scale Drivers have replaced the Development Mode of COCOMO 81. The first two Scale Drivers, Precedentedness and Development Flexibility actually describe much the same influences that the original Development Mode did [10].

C. Cost Drivers

COCOMO II has 17 cost drivers; you assess your project, development environment, and team to set each cost driver. The cost drivers are multiplicative factors that determine the effort required to complete your software project. For example, if your project will develop software that controls an airplane's flight, you would set the Required Software Reliability (RELY) cost driver to Very High. That rating corresponds to an effort multiplier of 1.26, meaning that your project will require 26% more effort than a typical software project. COCOMO II defines each of the cost drivers, and the Effort Multiplier associated with each rating [11].

D. COCOMO II Effort Equation

The COCOMO II model makes its estimates of required effort (measured in Person-Months -- PM) based primarily on your estimate of the software project's size (as measured in thousands of SLOC, KSLOC)):

$$\text{Effort} = 2.94 * \text{EAF} * (\text{KSLOC})\text{E} \qquad (2)$$

Where EAF Is the Effort Adjustment Factor derived from the Cost Drivers E Is an exponent derived from the five Scale Drivers.

As an example, a project with all Nominal Cost Drivers and Scale Drivers would have an EAF of 1.00 and exponent, E, of 1.0997. Assuming that the project is projected to consist of 8,000 source lines of code, COCOMO II estimates that 28.9 Person-Months of effort is required to complete it: Effort = $2.94 * (1.0) * (8)^{1.0997}$ = 28.9 Person-Months [8].

E. Effort Adjusment Factor

The Effort Adjustment Factor in the effort equation is simply the product of the effort multipliers corresponding to each of the cost drivers for your project.

For example, if your project is rated Very High for Complexity (effort multiplier of 1.34), and Low for Language & Tools Experience (effort multiplier of 1.09), and all of the

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

other cost drivers are rated to be Nominal (effort multiplier of 1.00), the EAF is the product of 1.34 and 1.09.

$$\text{Effort Adjustment Factor} = EAF = 1.34 * 1.09 = 1.46 \quad (3)$$

$$\text{Effort} = 2.94 * (1.46) * (8)^{1.0997} = 42.3 \text{ Person-Months} \quad (4)$$

F. COCOMO II Schedule Equation

The COCOMO II schedule equation predicts the number of months required to complete your software project. The duration of a project is based on the effort predicted by the effort equation [7]:

$$\text{Duration} = 3.67 * (\text{Effort})^{SE} \quad (5)$$

Where Effort Is the effort from the COCOMO II effort equation SE Is the schedule equation exponent derived from the five Scale Drivers Continuing the example, and substituting the exponent of 0.3179 that is calculated from the scale drivers, yields an estimate of just over a year, and an average staffing of between 3 and 4 people [7]:

$$\text{Duration} = 3.67 * (42.3)^{0.3179} = 12.1 \text{ months} \quad (6)$$

$$\text{Average staffing} = (42.3 \text{ Person-Months}) / (12.1 \text{ Months}) = 3.5 \text{ people} \quad (7)$$

G. The SCED Cost Driver

The COCOMO cost driver for Required Development Schedule (SCED) is unique, and requires a special explanation.

The SCED cost driver is used to account for the observation that a project developed on an accelerated schedule will require more effort than a project developed on its optimum schedule. A SCED rating of Very Low corresponds to an Effort Multiplier of 1.43 (in the COCOMO II.2000 model) and means that you intend to finish your project in 75% of the optimum schedule (as determined by a previous COCOMO estimate). Continuing the example used earlier, but assuming that SCED has a rating of Very Low, COCOMO produces these estimates:

$$\text{Duration} = 75\% * 12.1 \text{ Months} = 9.1 \text{ Months} \quad (8)$$

$$\text{Effort Adjustment Factor} = EAF = 1.34 * 1.09 * 1.43 = 2.09 \quad (9)$$

$$\text{Effort} = 2.94 * (2.09) * (8)^{1.0997} = 60.4 \text{ Person-Months} \quad (10)$$

$$\text{Average staffing} = (60.4 \text{ Person-Months}) / (9.1 \text{ Months}) = 6.7 \text{ people} \quad (11)$$

Notice that the calculation of duration isn't based directly on the effort (number of Person-Months) instead it's based on the schedule that would have been required for the project assuming it had been developed on the nominal schedule [7].

IV. DATA COLLECTION

In this section we introduce popular data sets of software cost and effort estimation. There are Cocomo, China, Desharnais, Finnish, Maxwell and Miyazaki Data sets. These datasets represent an interesting sample of industrial software projects collected from a single company or several software companies. The datasets cover a diversity of application domains and projects' characteristics. In particular, they differ for: observation number (from 38 to 499 projects); number and type of features (from 4 to 17 features, including a variety of features describing the software projects, such as number of developers involved in the project and their experience, technologies used, size in terms of Function Points [48], etc.); technical characteristics (software projects developed in different programming languages and for different application domains, ranging from telecommunications to commercial information systems); involved companies (the Desharnais dataset is within-company (WC), the others are cross-company (CC)); geographical locations (software projects coming from China, Canada, Finland). Furthermore all these datasets have been widely used in previous research work to evaluate effort estimation methods. Table 1 shows the datasets .

TABLE I. SHOWS THE DATASETS.

Data Set	Variables	No. of Records
COCOMO	Acap,pcap,aexp, Modp,tool,vexp,lexp, Sced,stor,data, time ,turn,virt, cplx, rely, Effort	63
China	Input ,Output, Inquiry, File, Interface, Effort	499
Desharnais	TeamExp, ManagerExp, Entities, Transactions , AdjustedFPs, Effort	77
Finnish	HW,AR , FP , CO, Effort	38
Miyazaki	SCRN, FORM, FILE Effort	48
Maxwell	SizeFP, Nlan, T01, T02, T03, T04, T05, T06, T07, T08, T09, T10, T11, T12, T13, T14, Effort	62

V. PROPOSED METHOD

In this paper we used cocomo model and its data for develop an artificial neural network approach for time estimation in software development process. The Constructive Cost Model (COCOMO) is an algorithmic software cost and time estimation model. The model uses a basic regression formula with parameters that are derived from historical project data and current as well as future project characteristics. as a model for estimating effort, cost, and schedule for software projects. It drew on a study of 63 projects at TRW Aerospace where Boehm was Director of Software Research and Technology. The study examined projects ranging in size from 2,000 to 100,000 lines of code, and programming languages ranging from assembly to PL/I. These projects were based on the

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

waterfall model of software development which was the prevalent software development process in 1981. The neural network model for effort estimation has 3 sections: the first section is data collection, second section: creating neural network model and finally: model evolution. We used COCOMO data set [7] for developing model. The COCOMO software cost model measures effort in calendar months of 152 hours (and includes development and management hours). COCOMO assumes that the effort grows more than linearly on software size; i.e. months=a* KSLOC^b*c. Here, "a" and "b" are domain-specific parameters; "KSLOC" is estimated directly or computed from a function point analysis; and "c" is the product of over a dozen "effort multipliers". I.e. months=a*(KSLOC^b)*(EM1* EM2 * EM3 * ...). The effort multipliers are as follows (Table II):

TABLE II. COCOMO FACTORS

State	factor	abbreviation
Increase these to Decrease effort	analysts capability	acap
	programmers capability	pcap
	application experience	aexp
	modern programing practices	modp
	use of software tools	tool
	virtual machine experience	vexp
	language experience	lexp
	schedule constraint	sced
decrease these to decrease effort	main memory constraint	stor
	data base size	data
	time constraint for cpu	time
	turnaround time	turn
	machine volatility	virt
	process complexity	cplx
	required software reliability	rely

In COCOMO I, the exponent on KSLOC was a single value ranging from 1.05 to 1.2. In COCOMO II, the exponent "b" was divided into a constant, plus the sum of five "scale factors" which modeled issues such as ``have we built this kind of system before?". The COCOMO~II effort multipliers are similar but COCOMO~II dropped one of the effort multiplier parameters; renamed some others; and added a few more (for "required level of reuse", "multiple-site Development", and "schedule pressure").

The effort multipliers fall into three groups: those that are positively correlated to more effort; those that are negatively correlated to more effort; and a third group containing just schedule information. In COCOMO~I, "sced" has a U-shaped correlation to effort; i.e. giving programmers either too much or too little time to develop a system can be detrimental.

VI. CREATING NEURAL NETWORK MODEL

In machine learning and cognitive science, artificial neural networks (ANNs) are a family of models inspired by biological neural networks (the central nervous systems of animals, in particular the brain) which are used to estimate or approximate functions that can depend on a large number of inputs and are generally unknown. Artificial neural networks are generally presented as systems of interconnected "neurons" which exchange messages between each other. The connections have numeric weights that can be tuned based on experience, making neural nets adaptive to inputs and capable of learning. Training a neural network model essentially means selecting one model from the set of allowed models that minimizes the cost criterion. Most of the algorithms used in training artificial neural networks employ some form of gradient descent, using backpropagation to compute the actual gradients. This is done by simply taking the derivative of the cost function with respect to the network parameters and then changing those parameters in a gradient-related direction. The backpropagation training algorithms are usually classified into three categories: steepest descent (with variable learning rate, with variable learning rate and momentum, resilient backpropagation), quasi-Newton and conjugate gradient [23]. Figure 1 shows the proposed neural network architecture. This network has 3 layers. The first layer has 16 neurons; the hidden layer has 15 neurons and finally in last layer has 1 neuron.

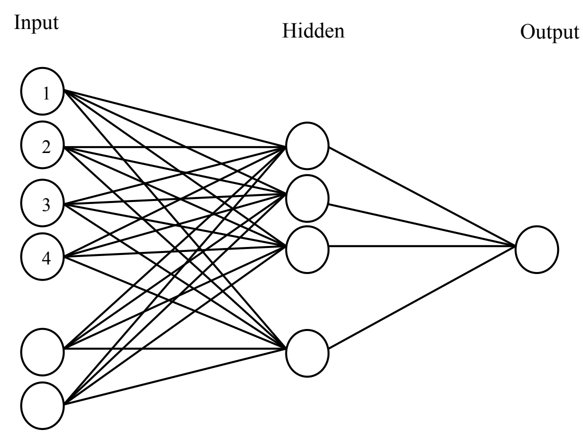


Figure 1. Neural Network Architecture

The model was then trained by using 37 (60% of dataset) data from the dataset, the remaining 13 (20% of the dataset) data and another 13 (20% of dataset) data were used to validate and test the model respectively. The data were randomly selected for all the cases by the neural network model.

For evaluation of proposed model we used MSE. In statistics, the mean squared error of an estimator measures the average of the squares of the errors or deviations, that is, the difference between the estimator and what is estimated. MSE is a risk function, corresponding to the expected value of the squared error loss or quadratic loss. The difference occurs because of randomness or because the estimator doesn't account for information that could produce a more accurate estimate. If YR is a vector of n predictions, and Y is the vector of observed values corresponding to the inputs to the function which generated the predictions, then the MSE of the predictor can be estimated by following:

$$MSE = \frac{1}{n}\sum_{1}^{n}(YR_i - Y_i)^2 \quad (12)$$

Figure 2 shows the best validation performance at epoch 2, this value is 0.0476.

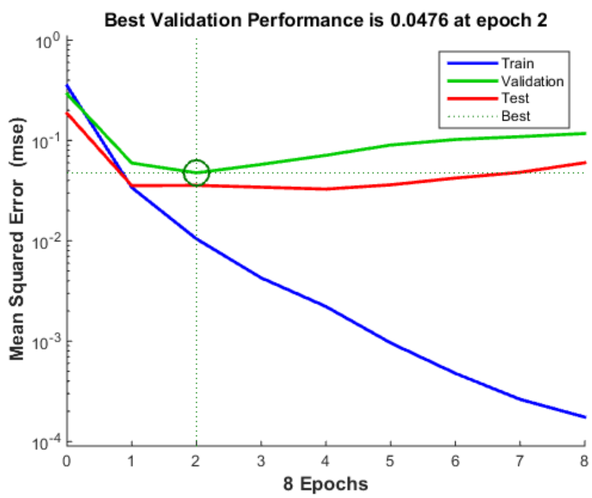


Figure 2. Best validation performance of network

Figure 3 shows the training state of network. At epoch 8 the gradient value is 0.02763 and Mu equal to 0.0000001 and validation checks equal to 6.

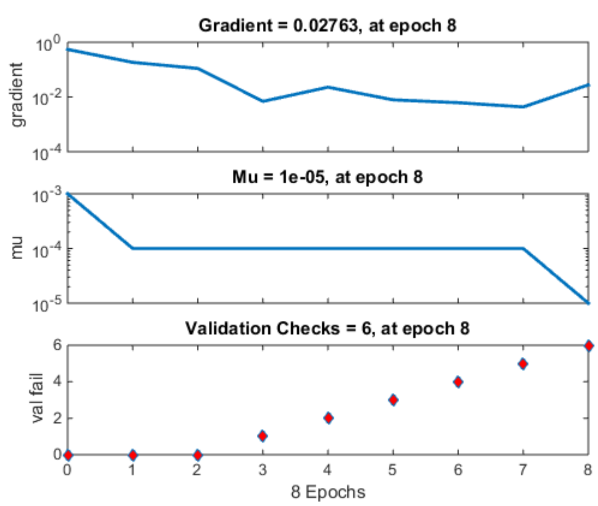


Figure 3. Training state of network

Figure 4 shows the error histagram of proposed model for software effort estimation. this figure shows the training, validation and test error of network in 20 epochs. The error is ERROR = Targets-outputs. Note that the most common ultimate goal of training is to minimize the error.

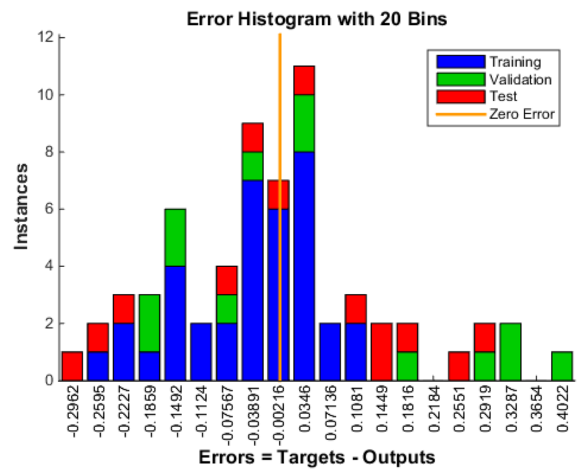


Figure 4. error histagram of proposed model

Table 3 shows the summarize of proposed neural network model for software effort estimation. Three experiments were conducted: The number of neurons 10, 15 and 20 were selected. When the number of neurons in hidden layer was 10, Number of training samples was 37, number of validation samples was 13 and number of testing samples was 13, the value of training, validation and testing MSE was 0.01860, 0.02097 and 0.5183 respectively and value of training, validation and testing R was 0.8231, 0.8552 and 0.6442 respectively. When the number of neurons in hidden layer was 15, Number of training samples was 37, number of validation samples was 13 and number of testing samples was 13, the value of training, validation and testing MSE was 0.02021, 0.05668 and 0.3013 respectively and value of training, validation and testing R was 0.8430, 0.5830and 0.5771 respectively. When the number of neurons in hidden layer was 20, Number of training samples was 37, number of validation samples was 13 and number of testing samples was 13, the value of training, validation and testing MSE was 0.01044, 0.0475 and 0.0375 respectively and value of training, validation and testing R was 0.9167, 0.7741 and 0.7410 respectively.

TABLE III. NEURAL NETWORK PERFORMANCE

No. of neurons in Hidden layer	State	Samples	MSE	R
10	Training	37	0.01860	0. 8231
	Validation	13	0.02097	0. 8552
	Testing	13	0.5183	0.6442
15	Training	37	0.02021	0.8430
	Validation	13	0.05668	0.5830
	Testing	13	0.3013	0.5771
20	Training	37	0.01044	0.9167
	Validation	13	0.0475	0.7741
	Testing	13	0.0375	0.7410

Figure 4 shows the regression plot of neural network. The training R is 0.91676, validation R equal to 0.77425, test R equal to 0.7411 and all R is 0.81554.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

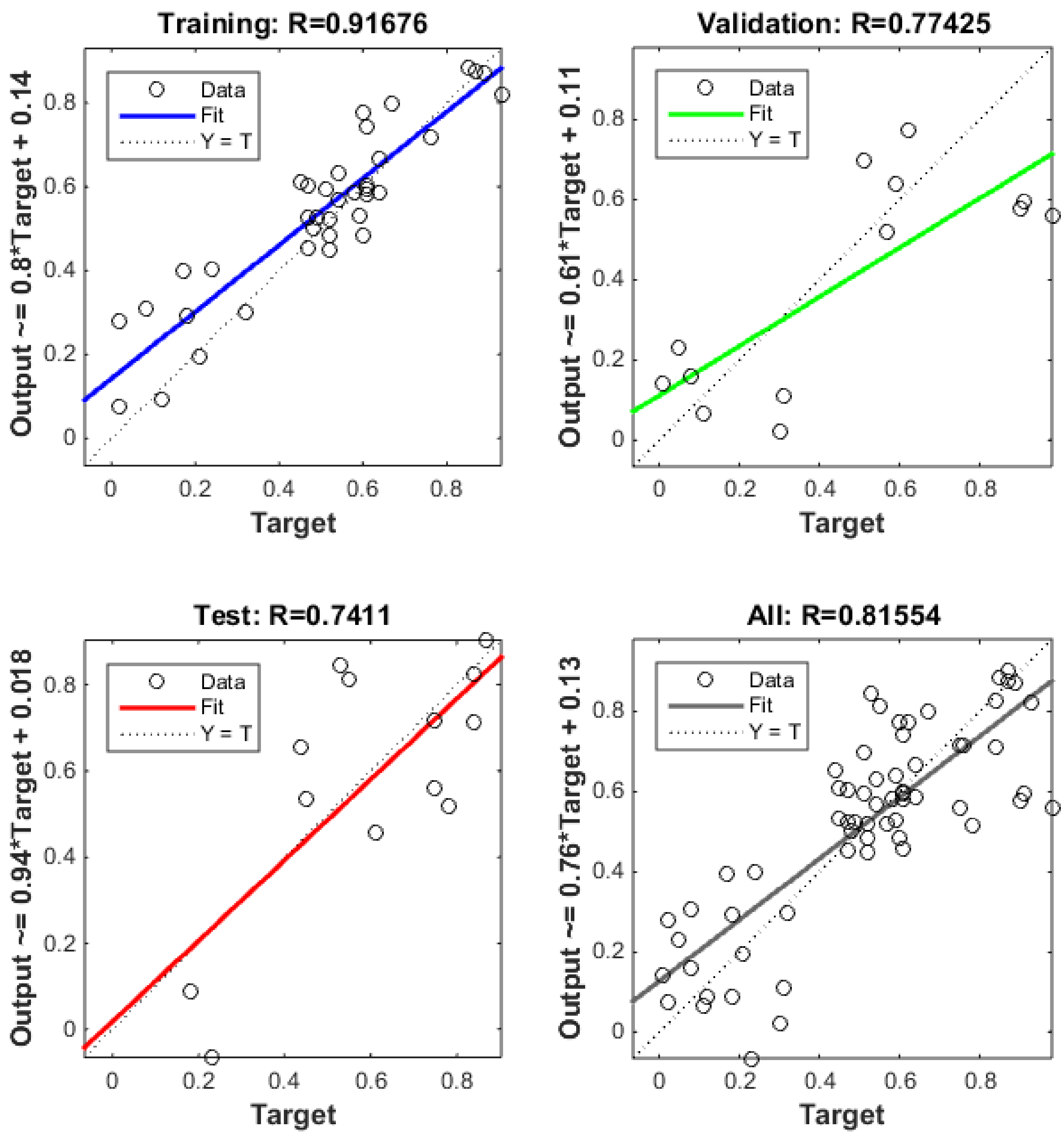


Figure5. The Regression Plot of Neural Network

VII. Conclusion

Software development effort estimation is the process of predicting the most realistic amount of effort required to develop or maintain software based on incomplete, uncertain and noisy input. Effort estimates may be used as input to project plans, iteration plans, budgets, and investment analyses, pricing processes and bidding rounds. Those responsible for effort estimations are usually the project managers. Depending on the chosen effort estimation method, they can estimate alone or with expert advice from developers, designers and testers. Other people that need most the effort estimation are project owners and sales. In this paper we proposed a neural network model for software effort estimation. This model has 3 layers. The train, validation and test data used are from COCOMO data set. Inputs and targets data randomly divided in train (60 %), validation (20%) and test (20%) group. Then we trained the network and test it. When the number of neurons in hidden layer was 20, Number of training samples was 37, number of validation samples was 13 and number of testing samples was 13, the value of training, validation and testing MSE was 0.01044, 0.0475 and 0.0375 respectively and value of training, validation and testing R was 0.9167, 0.7741 and 0.7410 respectively.

References

[1] Zahra, S.B. and Nazir, M., 2012. A Review of Comparison among Software Estimation Techniques. Journal of Information & Communication Technology, 5(1), pp.39-45.

[2] Balaji, N., Shivakumar, N. and Ananth, V.V., 2013. Software cost estimation using function point with non algorithmic approach. *Global Journal of Computer Science and Technology*, *13*(8).

[3] Bardsiri, A.K. and Hashemi, S.M., 2014. Software Effort Estimation: A Survey of Well-known Approaches. *International Journal of Computer Science Engineering (IJCSE)*, *3*(1), pp.46-50.

[4] Han, W.J., Lu, T.B., Zhang, X.Y. and Jiang, L.X., 2013. A new estimation model for small organic software project. *Journal of Software*, *8*(9), pp.2218-2222.

[5] Hamdy, A., 2014. Genetic Fuzzy System for Enhancing Software Estimation Models. *International Journal of Modeling and Optimization*, *4*(3), p.227.

[6] B. Boehm, Software Engineering Economics. Prentice Hall, 1981.

[7] B. Boehm, R. Madachy, and B. Steece, Software Cost Estimation with Cocomo II. Prentice Hall, 2000.

[8] L.H. Putnam, "A General Empirical Solution to the Macro Software Sizing and Estimation Problem," IEEE Trans. Software Eng., vol. 4, no. 4, pp. 345-361, July 1978.

[9] A.J. Albrecht and J.E. Gaffney, "Software Function, Source Lines of Code, and Development Effort Prediction: A Software Science alidation," IEEE Trans. Software Eng., vol. 9, no. 6, pp. 639-648, Nov. 1983.

[10] C.F. Kemerer, "An Empirical Validation of Software Cost Estimation Models," Comm. ACM, vol. 30, no. 5, pp. 416-429, 1987.

[11] A.F. Sheta, "Estimation of the COCOMO Model Parameters Using Genetic Algorithms for NASA Software Projects", Journal of Computer Science, Vol. 2, No. 2, pp.118-123, (2006).

[12] A. Mittal, K. Parkash, H. Mittal, "Software Cost Estimation Using Fuzzy Logic", ACM SIGSOFT Software Engineering, Vol. 35, No. 1, pp. 1-7, (2010).

[13] Prasad Reddy P.V.G.D, Hari CH.V.M.K, S. Rao.T, "Multi Objective Particle Swarm Optimization for Software Cost Estimation", International Journal of Computer Applications, Vol. 32, No.3, pp. 13-17, (2011).

[14] Ziauddin, Sh. K. Tipu, K. Zaman, Sh. Zia, "Software Cost Estimation Using Soft Computing Techniques", Advances in Information Technology and Management (AITM),Vol. 2, No. 1, pp. 233-238, (2012).

[15] B.K. Singh, A. K. Misra, "Software Effort Estimation by Genetic Algorithm Tuned Parameters of Modified Constructive Cost Model for NASA Software Projects", International Journal of Computer Applications, Vol. 59, No.9, pp. 22-26, (December 2012).

[16] L. Minku and X. Yao. Ensembles and locality: Insight on improving software effort estimation. IST, 55(8):1512–1528, 2013.

[17] Y. Kultur, B. Turhan, and A. Bener. Ensemble of neural networks with associative memory (ENNA) for estimating software development costs. Knowledge-Based Systems, 22:395–402, 2009.

[18] Jenkins, A.M., J.D. Naumann, and J.C. Wetherbe, *Empirical Investigation of Systems Development Practices and Results.* Information & Management, 1984. **7**: p. 73-82

[19] Bergeron, F. and J.-Y. St-Arnaud, Estimation of Information Systems Development Efforts: A Pilot Study. Information & Management, 1992. 22: p. 239-254.

[20] Heemstra, F.J., *Software cost estimation.* Information and Software Technology, 1992. **34**(10): p. 627-639.

[21] Heemstra, F.J. and R.J. Kusters, *Function point analysis: Evaluation of a software cost estimation model.* European Journal of Information Systems, 1991. **1**(4): p. 223-237.

[22] Heemstra, F.J., Software cost estimation. Information and Software Technology, 1992. 34(10): p. 627-639.

[23] Ge, S.S., Hang, C.C., Lee, T.H. and Zhang, T., 2013. Stable adaptive neural network control (Vol. 13). Springer Science & Business Media.

Authors Profile

Nesa Khandoozi GholiAbad,

She received her Bachelors of Software Enginnering from Islamic Azad University of Shirvan, Iran. Now, She is a master student of Software Enginnering in Islamic Azad University of Gorgan, Iran. Her research interest include software engineering, neural networks, fuzzy sets and cloud computing.

Sanaz Khandoozi GholiAbad,

She received her Bachelors of Software Enginnering from Islamic Azad University of Lahijan, Iran. Now, She is a master student of Software Enginnering in Islamic Azad University of Sari, Iran. Her research interest include software engineering, neural networks, fuzzy sets and cloud computing.

An Efficient Approach for Digital Image Splicing Detection Using Adaptive SVM

Gurleen Kaur(Research Scholar)
Department of Computer Science and Engineering Baba Banda Singh Bahadur Engg. College Fatehgarh Sahib, India

Baljit Singh Khehra(Associate Professor)
Department of Computer Science and Engineering Baba Banda Singh Bahadur Engg. College Fatehgarh Sahib, India

***Abstract*—Forgery detection is the most important task in our national judicial system and criminal investigation procedure. Today digital images have become powerful source of communication. With the advancement of technology, it becomes very easy to change the content of digital images. Due to which these images are no more taken as a proof of authenticity or legitimacy. In this paper, we deal with the widely used form of image tampering known as image composition(or image splicing).We demonstrate an effective algorithm to detect the spliced images based on illumination inconsistencies present in images. An adaptive support vector machine (a-SVM) is used to classify the given images as either genuine or forged.**

***Keywords*—Digital image forensic, forgery detection, image splicing, Adaptive SVM.**

I. INTRODUCTION

Digital images play an essential role in our daily life. With the advent of technology, these images have become powerful source of communication. Many areas such as, medical imaging, business, news, forensic investigation are seeking the benefit of these images. Nowadays, criminals of cyberspace are taking the advantage of powerful photo editing tools to falsify the information contained in digital images. Due to which, the trustworthiness of a digital image has become a challenging issue .Digital image forensic is new and growing field in the area of digital image investigation. This field has been working over past few years in order to revive some trust to digital images.

Digital image forensic can be classified into active forensics and passive forensics. In active forensic method, a watermark or digital signature is inserted at the time of recording, which would limit its usage by making it hardware or software specific. In contrast to these techniques, the passive methods are using the received image only for accessing its integrity. Hence, the passive forensic is more reliable for forgery detection. The most commonly used passive technique is image splicing (or image composition).In this process, two or more images are merged to form a single image. Figure 1 shows the steps to create a composite image.

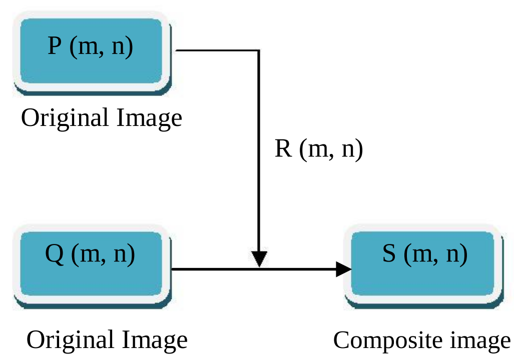


Figure 1: How to create a composite image?

An example of composite image is shown in figure 2.

(a) Original image (b) Composite image

Figure 2: Image splicing (or image composition)

Composite image forgery detection can be done in many ways. Various color based and geometry based methods have been proposed to detect the fake images based on illuminant color inconsistencies. In many cases, source of illumination is constant over a scene i.e. all objects present in the scene are illuminated by same light source. This property leads to the fact that illumination conditions on different objects present in the scene should be consistent. But when we create a fake

image, it become extremely hard to achieve these illumination conditions. Therefore, inconsistencies present in the scene are important indicator forgeries. Some of the existing techniques have certain kind of limitations. So there is a great need of problem specific forgery detection method based on color illuminance.

In this paper, we represent an effective approach towards forgery detection based on illumination inconsistencies. An automated forgery detection method is developed where the decision are taken by classifier to avoid the false judgment. Our focus is only on face images because objects of similar material can be effectively utilized to exploit the illuminant color inconsistencies. An adaptive support vector machine (a-SVM) is used to classify the original images and edited images.

II. RELATED WORK

In the field of illumination based forgery detection, various color-based and geometry-based methods are available. Here the deviation in light source position between specific objects present in the scene is captured by geometrical difference. Whereas, the disparity in the interactions between object color and light color is given by color-based methods.

Riess and Angelopoulou [19] suggested a new method to detect forgeries by creating illuminant maps. These illuminant maps are then used to detect the inconsistencies present in the scene. But this process involves manual processing of these maps.

Johnson and Farid [14] also suggested a method for splicing detection that makes use of specular highlights in the eyes. Specular highlights provide strong visual cue for the shape of object & location w.r.t light source in the scene. However, the major requirement of this method is that the eye of every individual must be available in high resolution.

Wu and Fang [2] indicate that illumination inconsistencies can be used to detect the forged images. They proceed their work by dividing an image into blocks. An illuminant color is locally estimated for each block and compared against a standard color to check whether an image is true or false. However, the accuracy of this method is very low.

Carvalho et al. [6] suggested a new method to detect false images. Here the illuminant color present in images is considered as major indicator of forgery. This method is based on machine learning approach. But this method is not fully automated and also there is a great need of machine leaning based illuminant estimators particularly for faces.

Bora et al. [16] proposed a novel method which is used to detect the composite images based on various color mismatches. The system works by estimating illuminant color from skin highlights. The dichromatic reflection model is for the said purpose. The illuminant color obtained is quantified using chromaticity coordinates. It is then matched against that of different persons in the composite image to detect the forgery. Here also there is a great need of alternative methods to find the illuminant color.

Neenu and Cheriyan [4] evinced a method that make use of illumination inconsistencies and resampling properties for detecting tampered images. This method is widely used in image forensics in order to check the authenticity of images.

III. PROPOSED METHOD

This section describes the techniques used in proposed forgery detection system.

A. *Collection of training samples to form the database:* The first and foremost step in the process of forgery detection is to collect and create true and false images. True images represent the original images collected from web and false images are the altered images that are created using some photo editing tools. These images are then used to train the adaptive support vector machine. After the completion of training process, a test set is used to check the performance of a-SVM classifier.

B. *Illuminant map creation:* After collecting the dataset, the next step is to partition these collected images into different sections having same color i.e. superpixels. Then is to calculate the illuminant color present in the images by using the pixels within each section. At the end, the extracted illuminant color is used for recoloring the entire section in order to create an illuminant map.
For the said purpose, the RGB image is converted into LUV coordinates to obtain so-called illuminant map. CIELUV color space is used to display color differences more conveniently. Here the luminancy is given by L* component and u*, v* define chrominancy. The color difference between two colors is given by ΔE i.e.
$$\Delta E = \sqrt{(L_2^* - L_1^*)^2 + (u_2^* - u_1^*)^2 + (v_2^* - v_1^2)^2}$$
Where ΔE represents the Euclidean distance of L*,u*,v* coordinates.

C. *Face extraction:* An automated face detector is used to create bounding boxes around the faces present in the image. There is no need of human expertise in the process of face detection. Also we are limiting our detector to skin, and more specific to faces in order to classify the illumination on a pair of faces as either consistent or inconsistent.

D. *Feature extraction:* Feature extraction is a special form of dimensionality reduction. The major aim of this process is to obtain the most relevant data in order to perform the desired task in a low dimensionality space. We extract the various gradient based and texture based features from all the faces present in the images.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

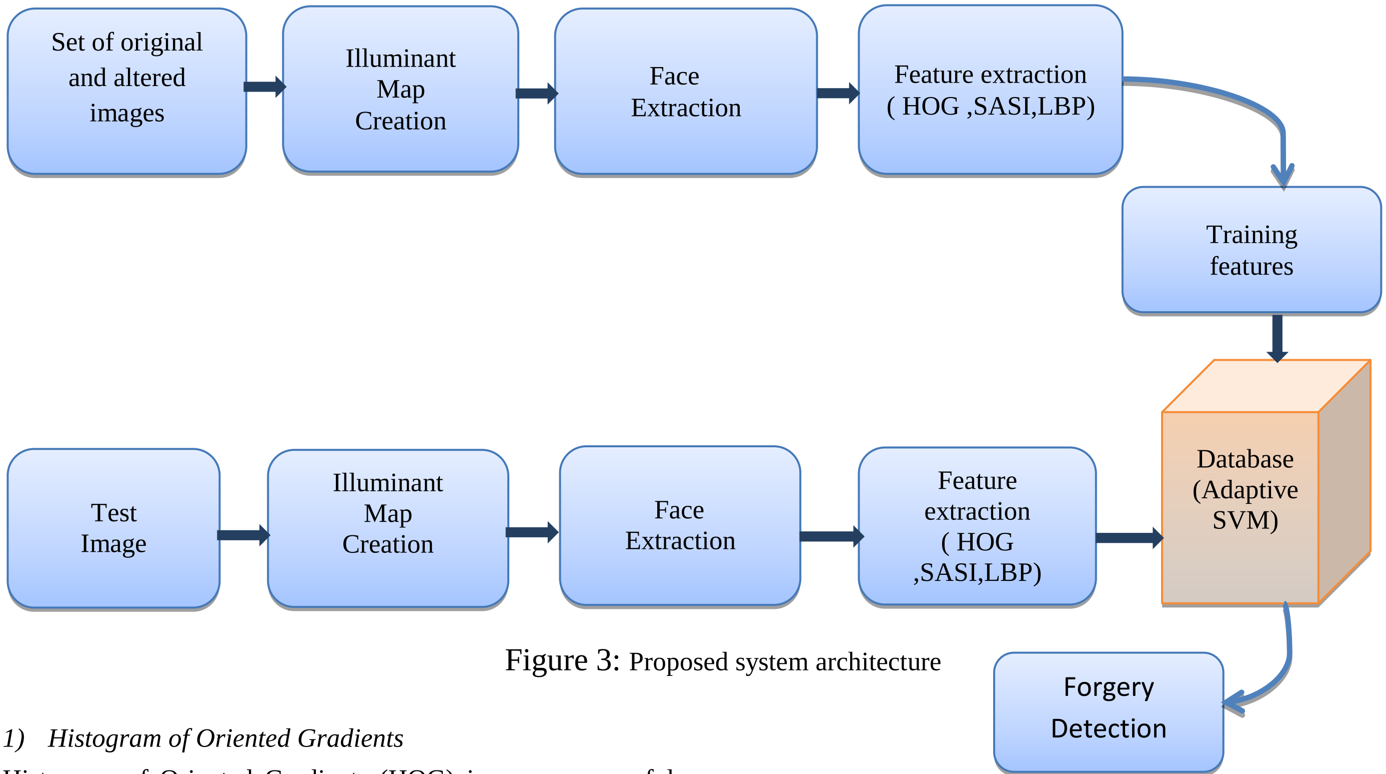


Figure 3: Proposed system architecture

1) *Histogram of Oriented Gradients*

Histogram of Oriented Gradients (HOG) is a very powerful feature extractor used in the area of image processing. It generally focuses on detecting the shape of structures present in images. For this purpose, it tries to capture the gradient information from these images. The implementation of HOG algorithm is given below:

Step 1. The first step is to partition the image into small cells usually of size 8x8 pixels.

Step 2. Each pixel present in the cell corresponds to gradient orientation bin. Also there are fixed number of gradient orientation bins in each cell. Separate each cell into angular bins according to gradient orientation.

Step 3. Calculate the weighted gradient from these angular bins.

Step 4. The next step is to combine these cells in order to form a block which is usually of size 4x4 cells.

Step 5. Normalize the histogram in accordance with their energy over blocks. The set of normalized histogram represent the block histogram and these blocks represent the feature descriptor.

HOG is very popular and widely used because of its ability to remain constant to various photometric and geometric changes.

2) *Statistical Analysis of Structural Information:*

Statistical Analysis of Structural Information (SASI) given by Yarman-Vural and Carkacioglu is used to extract texture information from illuminant maps. SASI is more advantageous because of its capability to identify the similar textures.

SASI is a generic descriptor based on the autocorrelation of horizontal, vertical and diagonal pixel lines over an image at different scales. Autocorrelation is computed using a specific fixed orientation, scale and shift. The mean value and standard deviation is calculated for all pixel values which yields a two feature dimensions. This vector is normalized by subtracting its mean value and dividing it by its standard deviation. The general algorithm of SASI given by Yarman-Vural and Carkacioglu [17] is as follows:

Step 1. Select the neighborhood system , where d is the order of neighborhood system.

Step 2. The next step is to choose the sizes S of clique window.

2.1 Calculate the lag vector v (k. l) used for each clique window.

Step 3. For each clique window W

3.1 For each lag vector v (k.l)

3.1.1 For each pixel

3.1.1.1 Define clique window W

3.1.1.2 Calculate r(k,l)

3.1.2 Calculate mean value and standard deviation of r(k,l)

Step 4. Construct vector and normalized vector.

3) *Linear Binary Pattern*

Linear Binary Pattern (LBP) is a type of visual descriptor used for classification in computer vision. It is also very powerful feature extractor for texture classification. When it is combined with histogram of oriented gradients, improves the detection performance to considerable amount. The algorithm for implementing LBP is as follows:

Step 1. Select the window and divide it into cell. Each cell may contain 16x16 pixels.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Step 2. Compare the each pixel present in the cell with its 8 neighbors in clockwise or counter clockwise direction.

Step 3. Set the pixel value to '0' if its value is greater than neighborhood pixel, otherwise write '1'.This method provides a 8-digit binary number.

Step 4. The next step is to compute the histogram for each cell.

Step 5. Normalize the histogram if necessary.

Step 6. Integrate these histograms which subsequently provide a feature vector for entire window. This feature vector can now be processed by using support vector machines.

E. Classification: In this phase, we classify the pair of faces as consistent or inconsistent. In other words, if all the faces present in the image are constantly illuminated by same light source then label the image as original otherwise forged. The various features extracted by HOG, SASI and LBP are given as input to the classifier. Here we use adaptive support vector machine in order to classify the data based on above extracted features.

Adaptive Support Vector Machine (A-SVM): The general idea behind adaptive SVMs is to adapt one or more existing classifiers for a new dataset that has limited labeled examples. The key problem is to choose an effective classifier for transformation. This problem can be solved by determining the performance of each existing classifier on the dispersedly labeled new dataset.

Here we consider a general classification problem in case of a primary dataset D^p that has a limited number of labeled examples represented by D_l^p and very large number of unlabeled examples represented as D_u^p. Thus whole dataset can be shown as:

$$D_p = D_l^p \cup D_u^p$$

Apart from this, there are also one or more secondary datasets represented as D_1^s ,.....,D_M^s .All these datasets are fully labelled and follow a different distribution than primary dataset. The secondary datasets are classified by using a secondary classifier f_k^s .When we are trying classify D^p (the primary dataset) with the help of f_k^s (the secondary classifier),it may not provide good results due to mismatched distributions. On the other hand, if we are choosing a new classifier learned with very few labelled examples in may also not suitable. To avoid these problems, it is required to use both the knowledge in secondary data and labelled primary examples for building an advanced classifier in order to classify the whole primary dataset.

The primary labelled dataset can be represented as:

$$D_l^p=\{(x_i - y_i)\}_{i=1}^N$$

Where x_i is the data vector and y_i is its binary label such that $y_i \in$ {-1, +1}.Also we set the value of first element of each dataset as constant 1 for the sake of notational simplicity such that $x_i \in R^{d+1}$, where d represent the number of features. In subsequent addition, we represent the fully labeled secondary dataset as follows:

$$D_k^s = \{(x_i^k, y_i^k)\}_{i=1}^k$$

Where $x_i^k \in R^{d+1}$ and $y_i^k \in$ (-1,+1).We consider a secondary classifier $f_k^s(x)$ that has been trained from each secondary dataset D_k^s , which predicts the data label through sign of its decision function i.e. $\hat{y} = f_k^s(x)$.

These secondary classifiers can be trained using any classification algorithm, such as decision tree, artificial neural network, naive bayes and support vector machines etc.

The major goal of this research is to learn a classifier f(x) that can accurately classify the primary dataset which is our original dataset. For this purpose, we use adaptive SVMs that can adapt the combination of multiple secondary classifiers $f_1^s(x)$,...., $f_M^s(x)$ to a new classifier $f(x)$ based on labelled examples D_l^p. The key idea behind this approach is to propose an effective adaptive SVM model that leverages the multiple secondary classifiers in order to adapt them to a new classifier $f(x)$.This adaptive classifier can be obtained by using some notations of standard SVMs.

In case of standard linear SVM, the label of a data vector x is determined by the sign of a linear decision function as $f(x) = w^T x$ such that w $\in R^{d+1}$ are the model parameters. While in case of non linear classification problem, this can be achieved by the use of "kernel-trick". In this case each data vector x is projected into a feature vector Φ(x) with the help of a feature map Φ such that:

$$f(x) = w^T \Phi(x) \tag{1}$$

Here the form of decision boundary is determined by the kernel function $K(x, x')$ = <*Φ*(x), (x')>, which defines the inner product of two projected feature vectors.

The training of standard SVM from $D_l^p = \{(x_i - y_i)\}_{i=1}^N$, introduces an optimization problem given as:

$$min_w \frac{1}{2} ||w||^2 + C\sum_{i=1}^N \xi_i \tag{2}$$

Such that $\xi_i \geq 0$, $y_i w^T (x_i) \geq 1 - \xi_i$, $\forall (x_i, y_i) \in D_l^p$

(i) $\sum_{i=1}^N \xi_i$ measures the total classification error.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

(ii) w in the given term || || represents a regularizer that is inversely proportional to the margin between training examples of two classes.

Our aim is to find a decision boundary that achieves a small classification. By using above notations we can easily represent an adaptive classifier in the following manner:

$$f(x) = \sum_{k=1}^{M} t_k f_k^s(x) + \Delta f(x) \tag{3}$$

(i) Where $t_k \in (0,1)$, is the weight of each secondary classifier $f_k^s(x)$ such that $\sum_{k=1}^{N} t_k = 1$, represent the sum of weights is going to be 1.

(ii) Whereas $\Delta f(x)$ represent a "delta function" used for adapting the secondary classifier trained from secondary data D^s to a new classifier $f(x)$ in order to classify primary data . By using equation (1), we can rewrite the equation (3) as:

$$f(x) = \sum_{k=1}^{M} t_k f_k^s(x) + w^T \Phi(x) \tag{4}$$

The objective function of adaptive SVM can be modeled as:

$$min_w \frac{1}{2} ||w||^2 + C\sum_{i=1}^{N} \xi_i$$

Such that $\xi_i \geq 0$, $y_i \sum_{k=1}^{M} t_k f_k^s(x_i) + y_i w^T(x_i) \geq 1 - \xi_i$.

(i) Similar to equation (2), $\sum_i \xi_i$ measures the total classification error of adapted classifier $f(x)$.

(ii) We represents the regularizer having same form as in equation (2), but with different meaning because here w are the linear parameters of $\Delta f(x)$ only. The regularizer supports the 'Δ' function close to '0' such that $\Delta f(x) = 0$, which follows that new classification function, is close to secondary classifier f_k^s .

(iii) However, C represents the cost function that should be s all for an "effective" secondary classifier & vice versa.

Hence the above objective function looks for a new decision boundary that must be close to the boundary of secondary classifier and it also must be able to accurately distinguish the labeled examples in labeled primary dataset (D_l^p) .

IV. EXPERIMENTAL RESULTS

In this experiment, we use an adaptive support vector machine for the classification .This system is implemented using Matlab 2014a.For the purpose of detecting forgery, a set of original and altered images are given as input to the classifier. The performance of this proposed is evaluated by comparing the results with existing forgery detection system.

1) *Composite image forgery detection dataset:* In order to carry out the process of forgery detection, a set of 200 images has been selected. Out of these, 100 original images are taken from Pinterest and the other 100 forged images are created using Photoshop. These images are further shown to 20 human observers with normal color vision. They are then asked to label these images as either genuine or fake. Based upon their decision, we evaluate the performance of our system.

2) *Performance evaluation:* Based upon human decision, we check the effectiveness of our system. Here we calculate the true positive rate (TPR) and false positive rate(FPR) for better accuracy.

$$TPR = \frac{Number\ of\ images\ detected\ as\ forged\ which\ are\ actually\ forged}{Total\ number\ of\ forged\ images}$$

$$FPR = \frac{Number\ of\ images\ detected\ as\ foged\ which\ are\ authentic}{Total\ number\ of\ authentic\ images}$$

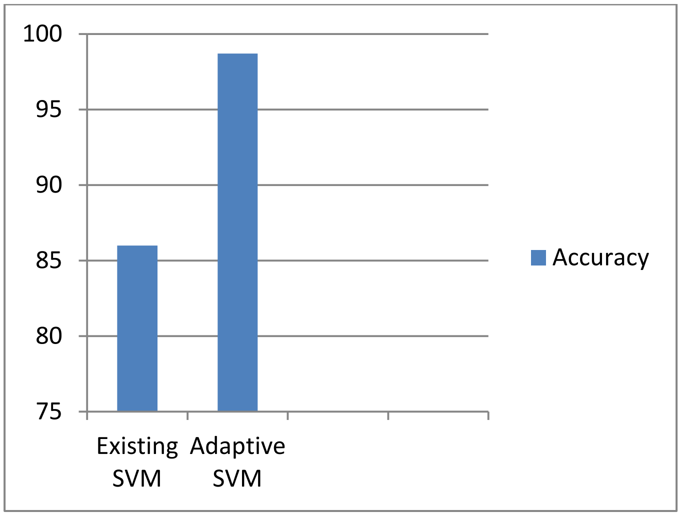
We train our system with the help of adaptive SVM. In order to detect the forgery based on color illuminance, HOG, SASI and LBP feature extractors are used for extracting illuminant features of an image. With the help of these features, we train the adaptive SVM to detect the forgeries present in digital images. The major advantage of adaptive SVM is its ability to adapt one or more existing classifiers for our primary dataset. We calculate the performance of proposed algorithm based on the accuracy in results with respect to existing system. It has been found that the existing system [6] for forgery detection performs well by yielding detection rates of 86% on a standard dataset. In existing system, they used SVM meta-fusion classifier in order to distinguish the original and altered images. However, by using an adaptive SVM for classification, the accuracy of system is increased by 98.7%.Also this work is fully automated and describes the authenticity of a given image.

S. No.	Method Used	Accuracy
1	T. Carvalho[6]	86%
2	Proposed system	98.7%

Table 1: Comparison of forgery detection techniques.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

The accuracy of both these methods is also shown in a graphical form as below:



V. CONCLUSION

An efficient method for detecting digital image forgeries is presented in this paper which is based on the concept of illumination inconsistencies. As we know that illumination inconsistencies present in the scene provide significant cues for detecting false image. Here the focus is to create an illuminant map from given images. These maps are then used to extract various edge based and texture based features. These features are further processed in training and testing phase of classifier. An adaptive support vector machine is used to classify whether the given image as genuine or forged. We can assume that our approach towards forgery detection, in addition to various forensic tools, may be effective in determining tampering detection.

REFERENCES

[1] G.K Birajdar and V.H Mankar, "Digital image forgery detection using passive techniques: a survey," *Digital Investigation*, vol. 10, no. 3, pp. 226-245, 2013.

[2] X. Wu and Z. Fang, "Image splicing detection using illuminant color inconsistency," in Proc. *IEEE Int. Conf. Multimedia Information*, pp. 600-603, 2011.

[3] S. Bianco and R. Schettini, "Color constancy using faces," in *IEEE Computer Vision and Pattern Recognition*, USA, 2012.

[4] H.U Neenu and J. Cheriyan, "Image forgery detection based on illumination inconsistencies and intrinsic resampling properties," in *IEEE Int. Conf. on Magnetics, Machines & Drives*, pp. 1-6, 2014.

[5] J.F.O Brien and H. Farid, "Exposing photomanipulation with inconsistent reflections," in *ACM Trans. Graphics*, vol. 31,no. 1, pp. 1-11, 2012.

[6] C. Riess,E. Angelopoulou T. Carvalho, "Exposing digital forgeries by illumination color classification," *IEEE Transaction on Inform. Forensics and Security*, vol. 8, no. 7, pp. 1182-1194, July 2013.

[7] M.A Qureshi and M. Deriche, "A bibliography of pixel-based blind image forgery detection techniques," *Signal Processing: Image communication*, vol. 39, no. 4, pp. 46-74, 2015.

[8] H. Farid, "A survey of image forgery detection," *IEEE Signal Processing Mag.*, vol. 26, no. 2, pp. 16-25, March 2009.

[9] S. Gholap and P.K Bora, "Illuminant color based image forensics," in *in Proc. of IEEE Regoin 10 Conference*, pp. 1-5, 2008.

[10] W. Fan, K. Wang ,F. Cayre and Z. Xiong, "3-D lighting based image forgery detection using shape-from-shading," in *in Proc. Eur. Signal processimg Conference(EUSIPCO)*, pp. 1777-1781, 2012.

[11] H. Farid, "A 3-D lighting and shadow analysis of the JFK zapruder film(frame 317)," Darmouth College, Technical Report, 2010.

[12] M. Johnson and H. Farid, "Exposing digital image forgeries in complex lighting environments," *IEEE Transaction Information Forensic Security*, vol. 3, no. 2, pp. 450-461, June 2007.

[13] E. Kee and H. Farid, " Exposing digital forgeries from 3-D lighting environments," in Proc. *IEEE International Workshop on Information Forensics and Security(WIFS).* Pp. 1-6, December 2010.

[14] M. Johnson and H. Farid, "Exposing digital image forgeries through specular highlights on eye," in Proc. International Workshop on Information Hiding, pp. 311-325, 2007.

[15] G.H Wang,Y. Bian,Z. Yu and Z. Zhang, "A novel model for splicing detection," in *in Proc. IEEE Conference on Bio-inspired Computing: Theories and Applications*, pp. 962-965, 2010.

[16] S. Gholap and K. Francis P.K Bora, "Illuminant color based image forensics using mismatch in human skin highlights," in *IEEE 20th National Conference on Communications*, pp. 1-4, 2014.

[17] A. Carkacioglu and F.T. Yarman-Vural, "Sasi: A generic texture descriptor for image retrieval," *Pattern Recognition*, vol. 36, no. 11, pp. 2615-2633, 2003.

[18] R. Yan and A.G. Hauptmann J. Yang, "Cross-domain vedio concept detection using adaptive SVMs," in *in Proc. ACM*, 2007.

[19] C. Riess and E. Angelopoulou, "Scene illumination as an indicator of image manipulation," Inf. Hiding, vol. 6387, pp. 66–80, 2010.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 6, June 2016

COMPARISON AND ANALYSIS OF IMAGE SPLICING DETECTION USING ARTIFICIAL NEURAL NETWORKS

Lovepreet Kaur
Department of CSE & IT
BBSBEC, Fatehgarh Sahib, India

Dr. Baljit Singh Khehra
Department of CSE & IT
BBSBEC, Fatehgarh Sahib, India

***Abstract*— Due to advancement in technology it is easy to modify the digital images and the discovery of modified images can be the difficult task as the images are the very powerful source of communication in every field. So, one of the major issue in today's world regarding digital images is the authenticity of given images. Therefore, digital image forgery detection is a growing research field with important implication for ensuring the credibility of digital images. In this research, we proposed a credible method to detect image splicing based on illuminant color. Artificial neural network techniques are implemented as a classifier to detect the tampered images. The results describe that artificial neural network is effective to detect tampered images.**

Keywords*— *Forgery Detection, Image splicing, Illuminant color, Artificial Neural network.

I. INTRODUCTION

The security of digital content involves the authenticity of digital images, digital information that is broadcast in the digital form. As in today's world the images can be easily manipulated with the help of various image retouching and transform applications and in recent years, these transforming tools have increased day by day. So, detection of an image is very important to determine which image is real or fake by forgeries.

Digital image tampering detection involves the detection of altered images to examine the fake and real images for the security purposes. Digital image tampering detection techniques based on two approaches:

- Active Approach: In a active approach, there is need of pre-processing operations to generate and embed any watermark. Digital watermarking and Digital signature are active approaches. Digital watermarking is a common example of the active approach.
- Passive Approach: In a passive approach, there is no need of pre-processing any digital signature to be generated or be embedding any watermark. Passive approaches can be pixel based, camera based and physics based. In this work we are focusing on splicing part of pixel based forgery which is the common form of alteration in Images. The Figure shows the various tampering detection techniques.

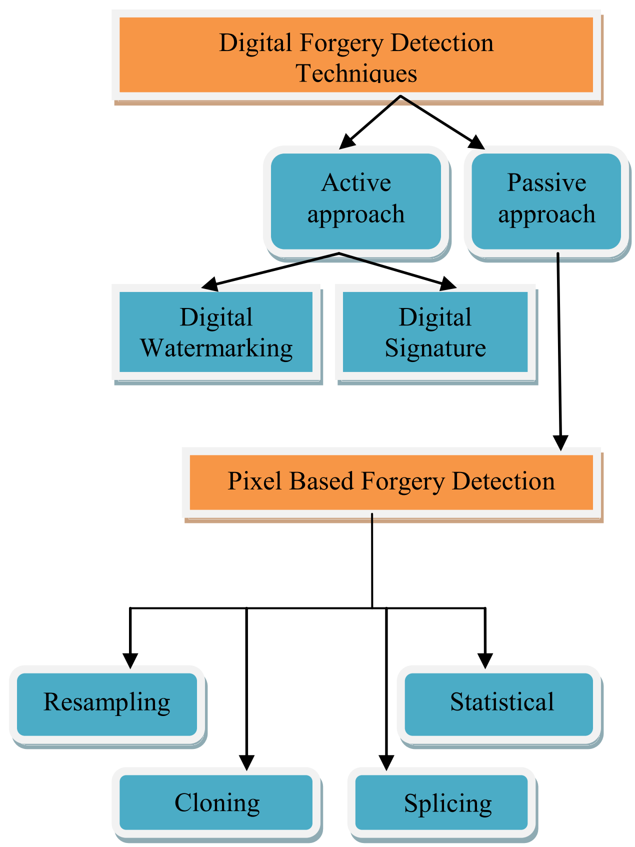


Figure 1: Digital Image Forgery Detection Techniques

In this paper, we introduce an automatic forgery detection system based on illuminant color for image splicing detection. Image splicing is very common for image manipulation. Image splicing is the simple process of cropping and pasting regions from the same and different images for the creation of tampered images. So, we apply three artificial neural networks as the classifier for the detection of tampered images. Experimental results show that Cascade forward back propagation neural network splicing detection algorithm is

effectual and authentic for the detection of tampered images. The figure shows the image splicing.

Fig (a) original image Fig (b) Tampered image

Figure 1.1: shows the image splicing.

II. RELATED WORK

Z. Zhang and G. Wang et al. [4] provided a system that is developed by measuring image quality metrics (IQMs) and squeezing some moment features. The model can compute statistical differences between original and altered images. In this work, (ANN) used as the classifier to detect the altered images. Experimental results show that the new splicing detection algorithm provides better accuracy in the detection of tampered images that is very important for security.

R.F and O. Khalifa et al. [5] proposed an algorithm based on fast Fourier transform and complex-valued neural network (FFT-CVNN) that can be used for watermarking medical images and the tamper detector was able to detect any forgery.

T Carvalho and C Riess et al. [3] recommended a forgery detection model based on the machine-learning approach and there is no need of any human interaction for the detection of tampered images. In this paper, the machine-learning approach used to edge and texture based features for the automatic detection of the tampered images. Here the classification is done by support vector machine. This model provides 86% accuracy in the detection of altered images.

Z. Moghaddasi et al. [21] recommended a model based on singular value decomposition (SVD).In this paper, detection of altered images done by merging singular value decomposition features and discrete cosine transform. Support vector machine was used to check the extracted features and detection of tampered image. (SVD+SVD-DCT) provides best detection rate compared to the individual methods SVD and SVD-DCT.

Z. Moghaddasi et al. (2015)[22] proposed an SVD- based image splicing detection method and tested in different spatial and frequency domains, discrete cosine transform (DCT), discrete wavelet transform(DWT), discrete Fourier transform (DFT). SVD-DCT has the best detection rate. SVD-DWT and SVD-DFT do not provide best results. Future Research is required to modify the SVD to improve the performance.

III. PROPOSED METHOD

We proposed a method for the detection of tampered images based on artificial neural networks. There are various steps that describe the detection of tampered images.

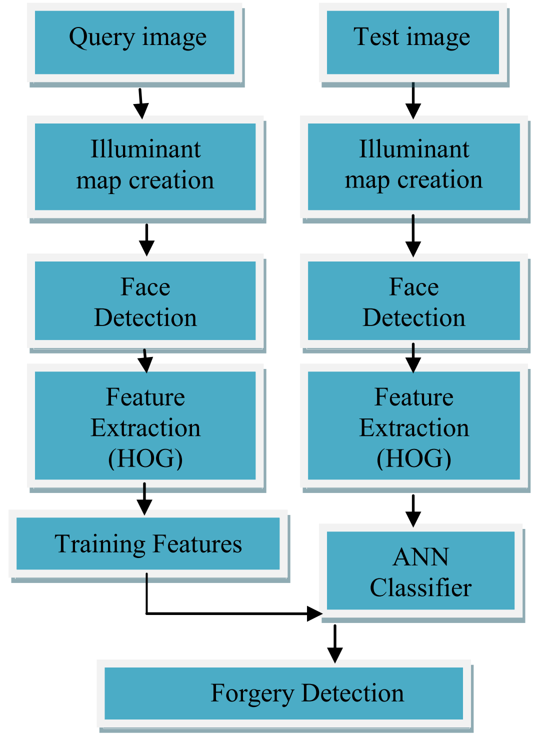


Figure 2: Overview of Forgery Detection method

A. *Illuminant Map Creation*: Illuminant map evaluates the color permanence and illuminant maps represent intermediary description. In this paper, for the creation of Illuminant map RGB image is converted into LUV coordinates and the L component describes the luminancy. The LUV coordinates represents the color differences more accurately and provide better results.

Figure 2.1 Represents the Illuminant map

B. *Face Extraction:* Face extraction is the primary step for forgery detection. In this system, face extraction

is fully automated and there is no human interaction. Detection of forged images will be done by extracting the faces that are present in the image with the help of automatic face detector. Face detectors create the bounding boxes over the each face of the image.

Figure 2.2 Face Detection Using Face Detector

C. *Feature Extraction*: Feature Extraction is useful in the case of large size images and feature extraction describes the effective sectors of the image. In this work, Features are extracted with the help of HOG (Histogram of Oriented Gradients) for the forgery detection. There are various Steps of HOG descriptor Algorithm described as:
- Split the image into small linked regions called cells.
- A histogram of gradient direction is computed for the pixels within each cell.
- The descriptor is the combination of histograms.
- Then for improved accuracy, the local histograms can be contrast-normalized with the help of intensity measured across the block. This normalization provides effective results in forgery detection.

D. *Classification:* Classification refers to the labeling of altered and real images. In the previous step, features that are extracted with the help of HOG are given as input to classifier to detect the forged images. In this work, we apply artificial neural network as the classifiers. Here, we apply feed-Forward neural network, Elman neural network and cascade forward backpropagation neural network for the classification of original and tampered images. Artificial neural network is a family of extensively parallel frameworks that are capable to generate efficient solutions in the case of incomplete data. So, the artificial neural network has the capability to solve complex engineering problems.
- Artificial neural network is a network that has capability to preserve its operations and a system with neurons, connections, and local memory. The artificial neuron measures activation and attribute output. Neuron is repressed of an integrator($\sum$) .
- The activity of neuron selected as a function of the activity of one or more neurons. The number of essential parameter s described in equation as :

$$O_i = f(\vec{x}, \overrightarrow{w_i}, \theta_i) \quad (1)$$

In this, $\vec{x}$ describes the vector of inputs from$x_1 \,..\, x_n$ and other neurons are distinguished by the set of weights $\overrightarrow{w_i}$ .The threshold θ_i represents the range of value.The activity of neuron (i) split into two parts.

Step 1. The first part describes the t is the scalar quantity that is expressed as:

$$net_{i=\sum_{j=1}^{N} w_{ij} x_{ij}} - \theta_i \quad (2)$$

$$net_i = \sqrt{\sum_{j=1}^{N} \left(w_{ij} - x_j\right)^2} - \theta_i \quad (3)$$

- The Equation (2) represents an innermost product between the input activity pattern vector and the row vector of weights.
- The Equation (3) characterized the Euclidean distance between input activity vector and its ith vector of weights matrix.

Step 2. The second part of neuron activity represents the activation function $o_{i=}f(net_i)$ and there is one additional representation of activation function as:

$$\theta_{i=}f(net_i) = \begin{cases} 1\, net_i \geq T \\ 0 net_i < T \end{cases}$$

The sigmoid function expressed as:

$$o_i = f(net_i) = \frac{1}{1 + e^{-\lambda net_i}}$$

λ controls the gradient of the output.

1) *Feed-Forward Neural network*:

Feed-Forward neural network is an artificial neural network technique that is used to detect the altered images for the security purposes. Feed-Forward neural network contains the hierarchy of processing units that are organized in a particular series of two more sets of neurons or layers. The various steps that describe the process of the feed-forward neural networks are:

Step 1. Select the training variable:

$$\{in_i^p, out_j^p : i = 1 \ldots ninputs, j = 1 \ldots noutputs, p = 1 \ldots n\ patterns\}$$

Step 2. After the process of training, create the network with n-inputs and n-outputs via connections with w_{ij} weights.

Step 3. Produce the particular inceptive weights.

Step 4. Select suitable error function E (w_{ij}) and learning rate η.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Step 5. For training variable P, apply the weight update $\Delta w_{ij} = -\eta \partial E^{(w_{ij})} / \partial w_{ij}$ to each weight w_{ij}.

Step 6. Repeat step 5. Until the error function will be decreased.

2) *Elman neural network*
Elman neural network can scrutinize as the feed-forward neural network. Elman neural network consists of input layer, output layer, hidden layer and context layer. This neural network has the capability to store the internal states. In this work, we apply this neural network for classification to detect the tampered images.

3) *Cascade forward back propagation feed forward neural network*
This neural network provides better accuracy than feed-forward and Elman neural network. The Steps on which the learning process of cascade forward back propagation neural network depends that describes as:

Step 1. Initialize the weights with small random values.

Step 2. For each combination (p_q, d_q) in the learning sample:

a. In this firstly, Propagate the entries p_q and then forward through the layers of the artificial neural networks:

$$a^0 = p_q; a^k = f^k(w^k a^{k-1}), k = 1, \dots, M$$

b. Back propagate the sensitivities through the layers of the neural network.

$$\delta^M = -2F{'}\mathrm{M}(n^M)(d_q - a^M);$$

$$\delta^k = F{'}^k(n^k)(w^{k+1})^T \delta^{k+1}, K = M-1, \dots 1$$

Step 3. Transform the biases and weights:

$$\Delta w^k = -\eta \delta^k (a^{K-1})^T, K = 1, \dots, M,$$
$$\Delta b^k = \eta \delta^k, K = 1, \dots, M,$$

Step 4. Stop after getting the results otherwise, procedure starts from step 2.

In this work, Cascade forward backpropagation neural network gives better accuracy than feed-forward and Elman neural network.

IV. EXPERIMENTAL RESULT

The tool we used to get the result is MATLAB 2014a.The proposed system based on ANN in which three models Feed-Forward neural network, ELMAN neural network and cascade-forward backpropagation neural network as classifiers to detect the forgeries in images and the classification shows the result after completion the whole process.

Digital image dataset

In order to perform the detection of tampered images, the dataset of 200 images is generated. Out of 200 images, 100 images are original that are downloaded from the pinterest and 100 are tampered. The tampered images are generated with the help of photo editing tools.

Performance evaluation

To perform the forgery detection, we use artificial neural networks feed-forward, Elman neural network, and cascade forward backpropagation neural network. The features are extracted with the help of HOG and given as input to classifiers. Classifiers took the HOG features as input and evaluate the performance. In this research, three classifiers are used for classification to detect the forged images. The cascade forward back propagation gives efficient results than the feed-forward; Elman neural network.Table 1.1 shows the output variables of the proposed work.

TABLE 1.Shows the accuracy for different classifiers

Accuracy of Classifiers		
Sr. No	Model Name	Model Accuracy
1.	Feed-forward neural network	46%
2.	ELMAN neural network	50%
3.	Cascade-forward Backpropagation neural network	97%

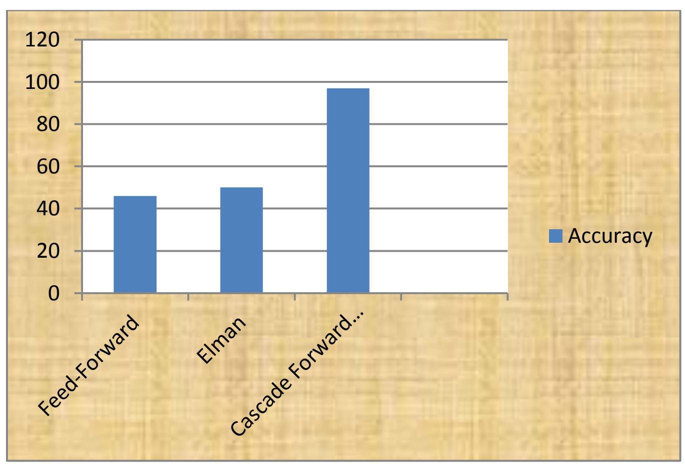


Figure 4.1 Graph shows the accuracy between Feed-Forward,Elman and Cascade-forward backpropagation neural networks and this shows that Cascade-forward backpropagation neural network provides improved and efficient output to detect altered images.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

V. CONCLUSION AND FUTURE WORK

Today, the use of digital images increasing rapidly so authentication of digital images is the most important part of security. In this work, we proposed a system for the detection of forged images using the illuminant color. The method is fully automated there is no human interaction and the three models Feed-Forward neural network, ELMAN neural network and cascade-forward backpropagation neural network that is used for classification to detect the real and tampered images based on the artificial neural network. In Future work, we will use other artificial neural networks for the better accuracy to improve the security by detecting the forged images with the help of artificial neural network classifiers.

References

[1] T.K. Huynh, T. Le-Tien, K.V. Huynh, S.C. Nguyen " A survey on image forgery detection techniques" IEEE RIVF International Conference on Computing & Communication Technologies Research, Innovation, and Vision for Future (RIVF), pp. 71-76, 2015.

[2] G.K. Birajdar and V.H. Mankar, "Digital image forgery detection using passive techniques: A survey," Digital investigation, vol. 10, no. 3, pp. 226-245, 2013.

[3] T. Carvalho, C. Riess , E. Angelopoulou, H. Pedrini and A. Rocha, "Exposing digital forgeries by illumination color classification", IEEE Transaction on Information Forensics and Security, vol. 8, no. 7, pp. 1182 – 1194, July 2013.

[4] Z. Zhang, G. Wang, Y. Bian and Z. Yu "A novel model for splicing detection," IEEE 5th International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA), pp.962-965, sept.2010.

[5] R.F Olanrewaju, O. Khalifa, A. Adulla, A.K.M Khedher " Detection of alternation in watermarked medical images using Fast Fourier transform and complex-valued neural network" IEEE 4th conference Mecatronics(ICOM), pp.1-6, may.2011.

[6] Z. Moghaddasi, H. A. Jalab, R.M. Noor,"SVD-Imagesplicing detection", International conference on Information Technology and multimedia (ICIMU), pp. 27-30, 2014.

[7] Z. Moghaddasi, H. A. Jalab, R.M. Noor, "A comparison study on SVD-based features in different transform for image splicing detection", IEEE International conference,pp. 13-14, 2015.

[8] M. Shabanifard, M. Akhaee and M. Shayesteh, "New method for forensics detection based on 2D-Histogram and Zernike moments, " IEEE Int. ISC Conference Information Security and Cryptology , pp. 151-155,2012.

[9] H.U.Neenu and J.Cheriyan "Image forgery detection based on illumination inconsistencies & intrinsic resampling properties," Emerging Research Areas: Magnetics, Machines and Drives (AICERA/ICMMD), pp.1-6, July.2014.

[10] C-M. Hsu. J-C. Lee and W-K. Chen "An efficient detection algorithm for copy-move forgery", 10th Asia Joint Conference on Information Security.,pp. 33-36,2015

[11] S. Wenchang; Z. Fei; Q. Bo; L. Bin, "Improving image copy-move forgery detection with particle swarm optimization techniques", China communications, IEEE journals and magazines. Vol. 13, pp. 139-149, 2016.

[12] L. Qingzhong and H. Andrew, "A new approach for JPEG resizes and image splicing detection", Proc. ACM multimedia and security workshop, pp. 43-8, 2009.

[13] S.D Lin and T. Wu " An integrated technique for splicing and copy-move for image detection ", IEEE International 4th conference on Image and Signal processing(CISP), Vol. 2, pp. 1086-1090, 2011.

[14] I. Amerini, R. Becareli, R. Caldelli, A.D Mastio " Splicing forgeries localization through the use of first digit features", IEEE International workshop on Information Forensics and security(WIFS),pp. 143-148, 2014.

[15] Y. Fan, P. Carre, C. Fernandez-Maloigne "Image splicing detection with local illumination estimation", IEEE conference on image processing (ICIP), pp. 2940-2944, 2015.

[16] K. Khuspe and V. Mane "Robust image forgery localization and recognition in copy-move using bag of features and SVM", IEEE Conference on Communication,Information Computing Technology (ICCICT), pp. 1-5,2015

[17] https://in.pinterest.com.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

(IJCSIS) International Journal of Computer Science and Information Security,
Vol. 14, No. 6, June 2016

Efficient random sampling statistical method to improve big data compression ratio and pattern matching techniques for compressed data

Dr. Nishad P M
Associate Professor
Department of Computer Applications
Mar Athanasios College for Advanced Studies Tiruvalla
Kerala, India

Syam Sankar
Assistant Professor
Department of Computer Science & Engineering
NSS College of Engineering Palakkad, Kerala, India,

Abstract- This paper surveys various possibilities for pattern matching in compressed big data volume. Although various compression standards are available for compressing data, entire volume decompression is compelled before pattern matching, this in turn leads to increase in computational complexity as well as the space complexity. Some compressions algorithms give better compression ratio, at the same time, they are inefficient in decompression required for pattern matching. This paper evaluates the possibilities of pattern matching after compression without decoding. Also this paper experiments and proposes how the random sampling and its statistics will help to make better compression ratio in big data. The another objective of this work is to investigate the possibilities of pattern matching in big data without decoding and some of the standards are suggested based on this study and survey.

***Keywords*- Compression, Encoding, Decoding, Big data, compression ratio, computational complexity, space complexity, random sampling.**

I. Introduction

Big data has become a buzzword in the fields of technical, business and other industries all over the world. According to IBM, 2.5 quintillion bytes of data are being generated every day. Data captured by social networking sites, real time data gathered by various sensors, data part coming under stock exchanges, data generated through smart phones, abundant uploaded videos in YouTube, data captured by online shopping sites and bank transactions, data derived from scientific researches like Large Hadron Collider etc. are the major resources of data explosion happening now a days. Gartner, an information technology research company, defines big data as high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation. This massive collection of data takes various forms, mostly in unstructured type thereby handling and analyzing through traditional database management tools is difficult. Big data infra structure demands high processing and performance analysis on data and a staunch support for real time responses. Nearly 40 zettabytes (ZB) of data will get generated by 2020 as per the estimations from Oracle. The storage, management and analysis of big data are some of the major challenges to be addressed very deeply and vastly. The tech world is trying to deal with them before they become an unmanageable problem. One of the key challenges large companies are facing today due to big data explosion is increasing demands of storage capacity and they spend large amount of money, time and effort every year in connection with satisfying storage. Here comes the word *data compression.* It means simplifying data from its original form to more concise representation. Compression algorithms attempt to reduce the size of the data so that it requires less disk space for storage and also transmitting fewer bytes of data across network, which is then faster and saves bandwidth capacity. The compressed form of data is mostly achieved by reducing redundancy. Compression algorithms take two forms: lossy and lossless. With lossless compression, the original data can be reconstructed completely from its decompressed format. On the other hand, lossy compression changes the originality of the data and the decompressed form is an approximate imitation of the original. It eliminates non essential things from the data volume. Images and videos are mainly compressed to smaller size by using lossy compression techniques. Run length encoding, delta encoding, dictionary encoding, entropy encoding, transformation encoding etc. are various encoding or compression schemes. It is difficult to search for particular patterns and retrieve information from compressed big data representation since it lacks the natural structure. A possible solution can be the usage of approach called decompress-then-search method. This approach is time and space consuming. So the solution for this crisis is meaningful search on compressed data sequence. This paper experiments and describes various possibilities of random sampling and its statistics to achieve better compression ratio as well as investigates the available possibilities of pattern matching on encoded data sequence. This paper also suggests data leafing for achieving better compression ratio using random sampling.

II. Literature review

梧瞁哥轥ern matching is an important problem in computer Although several methods for pattern matching are available, the existing pattern matching algorithm fails in compressed files so this leads to greatest issue like decompression before pattern matching. If the size of a compressed sequence is minimum, then decompression is advisable before pattern

matching even though it has an addition cost. When considering compressed big data sequence, the time required for both decompression as well as pattern matching is huge because of its size. To alleviate this problem, an effective pattern matching algorithm without decoding is required.

Boyer-Moore type algorithm proposed by Shibata et. al. was an initial method to perform compressed pattern matching. A right to left comparison is done when analyzing pattern with the text and stops when mismatch occurs. The algorithm works successfully for small patterns [1].

Pattern Matching in Z-Compressed Files, proposed by Amihood Amir, tried to find all the occurrences of a pattern in a compressed text in time proportion to the size of the compressed text [2] without performing decompression.

LZW Based Compressed Pattern Matching, by Tao Tao and Amar Mukherjee improved Amir's algorithm for pattern matching by including concepts like multi-pattern matching which uses Aho-Corasick algorithm. A faster implementation for so-called "simple patterns" was also proposed. [3].

G. Navarro and M. Raffinoth addressed the problem of pattern matching in Ziv-Lempel compressed text and also developed a hybrid compression technique which is between LZ77 and LZ78.They followed a general method for pattern matching when the text comes as a sequence of blocks, they [4].

Gonzalo Navarro found a solution to the problem of regular expression searching on compressed text and focused on LZ78 and LZW variants and also proved to search on compressed text twice as fast as decompressing plus searching. [5].

Kida et al initially compressed the text using LZW and applied a Shift-And approach to perform pattern matching in that compressed text. The Shift-And approach algorithm runs approximately 1.5 times faster than the decompression followed plus searching. They also proposed extensions to generalized pattern matching, to pattern matching with k mismatches and to the multiple pattern matching. [6].

Shibata et. al developed Byte Pair Encoding(BPE) for pattern matching in compressed text files. Substitution is the core technique of this method. To map every substitution made during compression, substitution tables are used. In the given text, frequently occurring pairs of characters are identified and it is replaced by a character that is none occurring in the text. Their experiments results shows that pattern matching using BPE compression is very faster than matching in the original text [7].

Farach and Thorup presents a LZ77 compressed matching algorithm to perform string matching in a compressed text without uncompressing it. For a given compressed string of size N, representing a text of size U, and a given pattern of size p the algorithm runs in time $O(N\log^2 U/N + p)$[8].

Takuya Kida et. al introduced a general framework to perform compressed pattern matching by following dictionary based compression and they finds all occurrences of a pattern in a text without decompress. ion[9].

Pawe l Gawrychowski, in his paper describes a method of performing pattern matching in Lempel-Ziv compressed strings and is an improved algorithm of the one which developed by Farach and Thorup[8] and a running time of $O(n \log (N/n) + m)$ was noticed.[10]

LeszekGasieniec and WojciechRytter describe almost-optimal pattern matching algorithms for compressed texts. For compression, they uses LZW. The algorithm runs in $O((n+m)\log(n+m))$ time on a single processor machine [11].

Matsumoto et. al developed a bit parallel approach for approximate string matching and achieved a time and space complexities of order $O(k^2n + km)$ and $O(k^2n)$ respectively for LZW compressed data [12].

Juha Kärkkäinen et. al presented a solution to the problem of approximate pattern matching over Ziv–Lempel compressed text. The solution can find the R occurrences of a pattern of length m allowing k errors over a text compressed by LZ78 or LZW into n blocks in O(kmn+R) worst-case time and $O(k^2n+R)$ average case time [13].

Burrows-Wheeler Transform (BWT) [14] compression algorithm processes a block of text as a single unit. A reversible transformation process which includes rotation, sorting and character extraction is applied to the block of characters to create a new block that contains the same characters. Then the new block is compressed by locally adaptive algorithms like move to front coding. Sorting is the key process of BWT compression. To improve compression ratio, an alternative alphabet ordering (a way of sorting) based on both heuristic and structured techniques was developed [15].

Run-Length Encoding (RLE) is one of the simplest compression technique in which the repeated symbols get replaced with a pair containing the length of the string and the symbol itself. It produces an output size two times more than the size of the input in worst case situation, which means the input with no repetition.

Huffman coding algorithm [16] is one of the successful methods of text compression. It uses a bottom approach for constructing Huffman code tree which in turn assigns shorter code words to more frequently occurring characters.

The family of Lempel Ziv algorithms [17] provided a basis for lossless data compression techniques. It follows a dictionary based approach and sliding window for achieving compression.

In order to reduce the computational complexity of LZW a novel approach is proposed using famous data mining technique called clustering. The algorithm is called index k nearest twin neighbor (IKNTN) clustering algorithm [18]. The computational complexity to find the availability of pattern in the dictionary is minimized in the proposed approach.

III. PROPOSED METHODOLOGY

Definitions:-B is big data, the size of B is |B|. B is fragmented in to fixed sized leaf i.e., $l_1, l_2, \ldots, l_n$, where $|l_1| = |l_i|$ and i= 2 to n. l_n indicates the last fragment or last leaf. $l_ir_1, l_ir_2, \ldots l_ir_n$ is used to represent various random samples in l_i.

the size of each r_i is fixed i.e., S or constant. The number of random sample taken from each leaf is C. A1, A2… An are the different algorithms used.

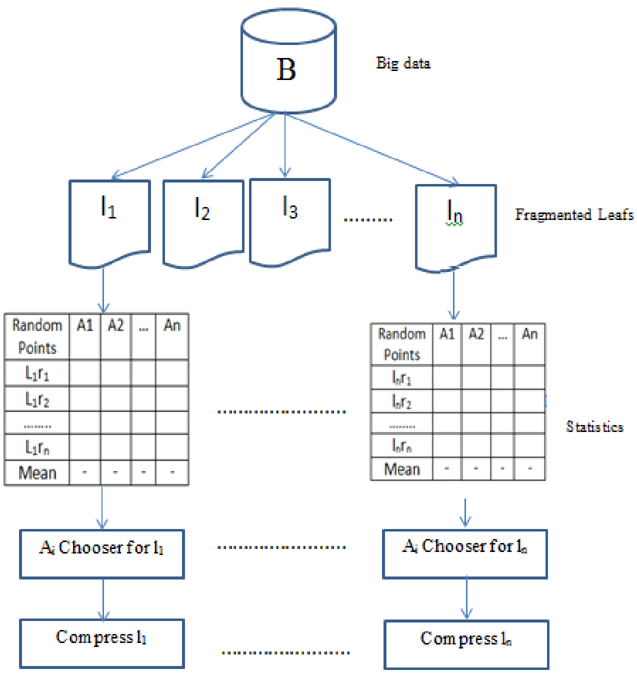


Figure-1 big data compression using proposed methodology

The proposed big data compression consist of three stages

1. Prepossessing
2. Algorithm selection
3. Compression

Prepossessing: - in this stage the huge volume of data is fragmented in to small units called fragmented leafs. i.e., the B is divided in to fixed sized smallest group namely l_i where i = 1 to n. This approach has several advantages. The additional over head of decompression is removed. For example if we need a specific portion of B after compression the entire volume must decompressed or decoded, so each time decoding entire volume lead to a huge overhead. Using the proposed approach only has to decode the corresponding fragmented leaf instead of entire volume. These increase the efficiency and reduce the computational complexity and remove the huge overhead during the encoding. The size of each fragmented leaf is in the same size, so no need of additional process required finding the leaf size also the re assembly is very easy when it required.

Algorithm selection:-The algorithm selection section stage has following steps

- ✓ Classification
- ✓ Fixed sized Random Point selection
- ✓ Test selected algorithm to check ability
- ✓ Build the static table
- ✓ Calculate the mean
- ✓ Select the algorithm based on the statistics

Classification: - in this stage the nature of data is tested and each l_i is classified based on its nature to get the better compression ratio. For example for the image preferably used the lossy algorithms and text type lossless is used. So based on the type and nature of the fragmented leafs the algorithms are selected from the list. i.e., A1, A2, …, An.

Fixed sized Random Point selection: - from each l_i, the l_ir_i, are selected. All the selected l_ir_i, are in the same size S.

Test selected algorithm to check ability:- Each l_ir_i, is compressed using the selected algorithm based on the previous stage.

Build the static table:- Compression ratio of each Ai is calculated for each l_ir_i, and a statistic table is maintained for every l_i shown in the fig-1

Calculate the mean:- Calculate the mean for each algorithm form recorded compression ratio

Select the algorithm based on the statistics:- based on the compression ratio the best A_i is selected for compressing l_i

Compression: - this is the final stage of this proposed approach Compress the li using the selected Ai.

The proposed approach has several advantages. Blindly selecting single algorithm for compressing the entire volume in B may not give best result. Each l_i in B has its own nature. For example some li may have low probability of symbols and high density. Some l_i may consist of high probability and low symbol density. In some both are in an average. So the same algorithm is not sufficient for all probability distribution. This is a disadvantage of traditional approach. This problem is completely solved using the proposed approach. Also some algorithms support the pattern matching in compressed files (for example LZW, BWT, RLE and etc.) so such compression technique is used for compressing l_i then encoding is not mandatory for searching pattern in the l_i. this will reduce the additional overhead in pattern matching or searching. The stages of proposed approach are shown in the fig-1.

IV. EXPERIMENTATION

To evaluate the proposed approach, several benchmark files are used. Each files in the list is treated as L_i and each L_i is tested against several compression algorithms (LZW, Huffman, Shannon Fano, adaptive Huffman coding and arithmetic coding). The results are maintained in the separate static tables (table-1, table-2 and table -3). After evaluating the results of statistical tables, the performance of LZW is optimal. It shows that an average compression ratio is 29.69%. The existing method calculates the mean compression ratio of all algorithms. After evaluating the mean, the LZW gives the optimal compression ratio. So all files are compressed with the LZW. The proposed approach effectively utilizes the property of statistical table and checks each L_i against each algorithm. Then selects the best algorithm for each L_i for compression. Red marked figures are the best compression ratio for each L_i in the statistical tables. The proposed approach shows 31.46% of Compression ratio.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Table-1: Statistical table for LZW and Huffman

File Name	Compress Ratio/ LZW	Compress Ratio/ Huffman
Example1. doc	55%	57%
Example2. doc	60%	66%
Example3. doc	42%	45%
Example4. doc	88%	76%
Example5. doc	70%	60%
Example6. doc	46%	53%
Example7. doc	38%	46%
Example8. doc	51%	55%
Example9. doc	62%	59%
Example10. doc	55%	57%
Pict3.bmp	87%	81%
Pict4.bmp	93%	80%
Pict5.bmp	68%	78%
Pict6.bmp	73%	73%
Inprise. gif	-43%	-9%
Baby. jpg	-35%	-1%
Cake. Jpg	-41%	-2%
Candles. jpg	-32%	-1%
Class. jpg	-16%	-3%
Earth. jpg	-38%	-5%

Table-2 Statistical table for LZW and Huffman

File Name	LZW %	HUF %	File Name	LZW %	HUF %
2.doc	75	76	5.bmp	80	71
3.doc	42	38	6.bmp	87	4
4.doc	37	63	1.tif	32	7
5.doc	48	37	2.tif	35	6
6.doc	52	50	3.tif	53	40
1.txt	35	35	4.tif	40	4
2.txt	45	37	5.tif	84	68
3.txt	72	33	6.tif	36	48
4.txt	48	37	1.gif	40	0
5.txt	52	38	2.gif	37	-1
6.txt	55	36	3.gif	41	-1
1.bmp	62	37	4.gif	41	-1
2.bmp	84	67	5.gif	42	-1
3.bmp	26	48	6.gif	38	0
4.bmp	26	3	__	__	__

Table-3: Statistical table for various algorithms

File name	RLE	Shannon Fano coding	Huffman coding	Adaptive Huffman coding	Arithmetic coding	LZW
Bib	-2	31	34	35	35	52
book1	-2	40	43	43	43	50
book2	-2	37	40	40	40	44
news	0	32	35	35	35	39
obj1	10	18	19	24	25	21
obj2	-1	19	21	21	24	-23
paper1	-1	33	36	37	38	43
paper2	-2	38	42	42	42	50
progc	-1	32	33	34	35	39
progl	3	36	39	40	41	51
progp	7	34	38	39	39	53
trans	1	27	30	30	31	47

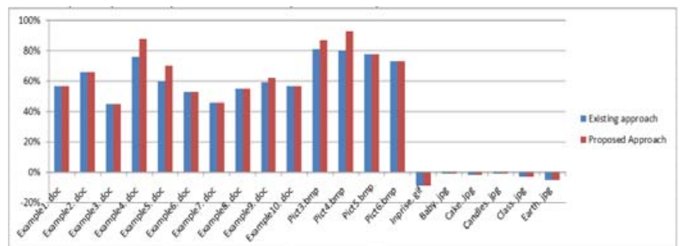


Figure 2: Performance analysis for statistical table 1

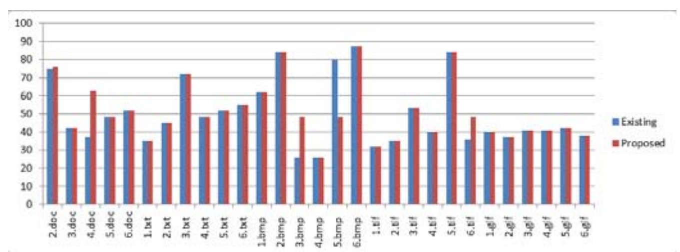


Figure 3: Performance analysis for statistical table 2

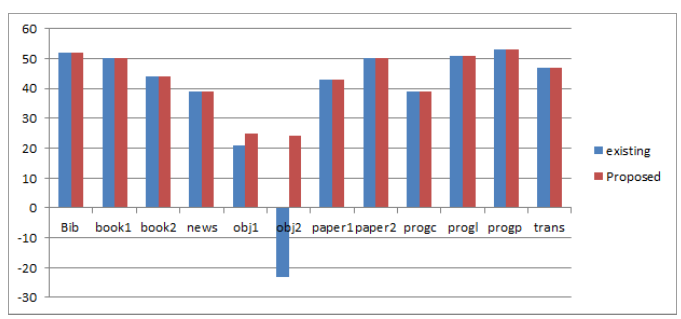


Figure 4: Performance analysis for statistical table 3

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

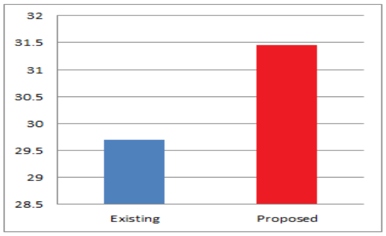


Figure 5: Overall Performance analysis

CONCLUSION

The proposed random sampling method chooses effective compression algorithm for each L_i and it gives best performance against the existing method. Unlike decompressing the entire data volume, the proposed approach decompresses the required L_i so that the computational complexity is reduced. Pattern matching in compressed L_i is comparatively easier if the leaf is compressed with best algorithm obtained through random sampling approach. The proposed approach results better compression ratio when comparing with existing method. The random sampling method can be extended by considering any data compression algorithms**.**

REFERENCES

[1] Shibata, Yusuke, et al. "A Boyer—Moore Type Algorithm for Compressed Pattern Matching." *Combinatorial Pattern Matching*. Springer Berlin Heidelberg, 2000

[2] Amir, Amihood, Gary Benson, and Martin Farach. "Let sleeping files lie: Pattern matching in Z-compressed files." *Journal of Computer and System Sciences* 52.2 (1996): 299-307

[3] Tao Tao, Amar Mukherjee, "Pattern Matching in LZW Compressed Files", IEEECS, 2005, pg 929-937

[4] Navarro, Gonzalo, and Mathieu Raffinot. "A general practical approach to pattern matching over Ziv-Lempel compressed text." *Combinatorial Pattern Matching*. Springer Berlin Heidelberg, 1999

[5] Navarro, Gonzalo. "Regular expression searching over Ziv-Lempel compressed text." *Combinatorial Pattern Matching*. Springer Berlin Heidelberg, 2001.

[6] Kida, Takuya, et al. "Shift-And approach to pattern matching in LZW compressed text." *Combinatorial Pattern Matching*. Springer Berlin Heidelberg, 1999.

[7] Shibata, Yusuke, et al. "Speeding up pattern matching by text compression."*Algorithms and Complexity*. Springer Berlin Heidelberg, 2000. 306-315.

[8] Martin Farach, MikkelThorup, "String Matching in Lempel-Ziv Compressed Strings", ACM, 1995, pg 703-71

[9] Takuya Kida, Yusuke Shibata, Masayuki Takeda, Ayumi Shinohara, SetsuoArikawa, "A unifying framework for Compressed Pattern Matching", IEEE, 1999, pg 89-96.

[10] Gawrychowski, Paweł. "Optimal pattern matching in LZW compressed strings." *ACM Transactions on Algorithms (TALG)* 9.3 (2013): 25

[11] LeszekGasieniec and WojciechRytter, "Almost optimal fully LZW-compressed pattern matching", Proceedings Data Compression Conference, 1999, pg 316-325

[12] Matsumoto, Tetsuya, et al. "Bit-parallel approach to approximate string matching in compressed texts." *String Processing and Information Retrieval, 2000. SPIRE 2000. Proceedings. Seventh International Symposium on*. IEEE, 2000.

[13] Kärkkäinen, Juha, Gonzalo Navarro, and Esko Ukkonen. "Approximate String Matching over Ziv—Lempel Compressed Text." *Combinatorial Pattern Matching*. Springer Berlin Heidelberg, 2000.

[14] M. Burrows and D. J. Wheeler. "A Block–sorting Lossless Data Compression Algorithm," SRC Research Report 124, Digital Systems Research Center, Palo Alto, CA, May 1994

[15] Higher Compression from the Burrows-Wheeler Transform by Modified Sorting Brenton Chapin Stephen R. Tate Dept. of Computer Science University of North Texas P. O. Box 311366 Denton, TX 76203–1366

[16] Huffman D.A., "A method for the construction of minimum-redundancy codes", Proceedings of the Institute of Radio Engineers, Vol. 40, No.9, pp. 1098–1101, 1952.

[17] S. Senthil and L. Robert, Text compression algorithms - a comparative study, ictact journal on communication technology, december 2011, volume: 02, issue: 04

[18] Nishad P. M (Feb. 2014) "A Novel Approach to Reduce Computational Complexity of Multiple Dictionary Lempel Ziv Welch(MDLZW) using Indexed K Nearest Twin Neighbour(IKNTN) Clustering and Binary Insertion Sort Algorithm." Ph.D thesis, Bharathiar University,Coimbathore

Dr. Nishad PM received his M.Sc.,M.Phil and Ph.D in Computer Science from Bharathiar University Coimbatore He got seven years experience in teaching and five years in research. He has published seventeen papers in national level/international conference/journals. He has presented three seminars at national Level. Now he is working as Associate Professor in Department of Computer Applications at Mar Athanasios College for Advanced Studies Tiruvalla (MACFAST), Kerala, India. He is a member of The International Society for Environmental information Sciences - [ISIES] –Canada and International Association of Engineers [IAENG]. His research interests are Data Compression, Computational complexity Theory, Cryptography Data Structures, Object Oriented Technology, Image Processing, and Data Mining

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

.

Syam Sankar received his B.Tech in Computer Science and Engineering and M.Tech in Computer and Information Science from College of Engineering Perumon under CUSAT, Kochi in 2013 and 2015 respectively. He is presently working as Assistant Professor in Department of Computer Science & Engineering at NSS College of Engineering Palakkad Kerala, India.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 6, June 2016

A New Dynamic Data Replication Algorithm to Improve Execution Time in Data Grid

Soheila Malmoli Abbasi
Department of Computer Engineering,
Khouzesan Science and Research Branch,
Islamic Azad University,
Ahvaz, Iran

Department of Computer,
Ahvaz Branch, Islamic Azad University,
Ahvaz, Iran

*Mohammadreza Noorimehr
Department of Computer Engineering,
Khouzesan Science and Research Branch,
Islamic Azad University,
Ahvaz, Iran

Department of Computer,
Ahvaz Branch, Islamic Azad University,
Ahvaz, Iran

***Abstract*— Data grids provide large-scale geographically distributed data resources for data intensive applications. These applications handle large data sets that need to be transferred and replicated among different grid sites so availability and efficient access are the most important factors affecting the performance. It is obvious that, managing the volume of data is very important. Data replication is an important technique to reduces data access time which improves the performance of the system by creating identical replicas of data files and distributing them on grid sites. In this paper, we propose a novel dynamic data replication strategy called DRPF (Dynamic Replication of Popular File), which is based on access history and file's popularity. As grid sites within a virtual organization(VO) have similar interest of files, the basic idea of DRPF is to improve locality in accesses through increasing the the number of replicas in the VO. DRPF first selects the popular files that are needed to be copied to other nodes, then tries to find the best places for placement of new replicas by taking into account parameters such as the number of demands per site for files and bandwidth between replication sites. The algorithm is simulated using a data grid simulator, OptorSim. The simulation results show that our proposed algorithm has better performance in comparison with other algorithms in terms of job execution time and effective network usage.**

Keywords-*Data grid; replication; popular file; placement***

I. INTRODUCTION

Today, huge amounts of data are generated around the world in many fields such as scientific and engineering applications that are shared among researchers globally for further studying. The management of the huge distributed and shared data resources efficiently around the wide area networks becomes a significant topic for both scientific research and commercial application[1].

Grid technology is the best solution to this kind of problem. One of the most important types of grids is data grid. Data Grid is the highlight in the development of the Grid technology, which can be treated as a suitable solution for high performance and data-intensive computing applications[2]. The most important point in this type of grid is the management of such large amounts of data so that easy availability and effective sharing of data could be guaranteed. Data replication is a key technique to manage large data in a distributed manner by its nature, we can achieve better performance (access time) by replicating data in geographically distributed data stores[3] so dynamic replication aims to maximize chances of data locality. In other words, improving the efficiency and reliability are two main reasons for the replication of data [4].

A data replication process involves creating identical copies of a file and placing them onto multiple sites so that they can be accessed simultaneously from various locations[5]. In such systems if one of the files fails to work and if it is impossible to access it, the system simply switches to other replicas of the file to prevent any disorder. In general, data replication algorithms can be divided into two groups: static and dynamic [6]. Static replications create replicas based on a set of predefined rules and require full knowledge of the workload. Therefore, these algorithms can not adapt themselves to network changes. In contrast, dynamic strategies adaptable to changes in users behavior and do data replication based on the actual network conditions and access patterns [7]. File access pattern analysis has always been employed as a powerful tool to design efficient dynamic data replication schemes [8][9].
There are three key issues in all the data replication algorithms as follows [10]:

- When should the replicas be created?
- Which files should be replicated?
- Where should the replicas be placed?

Depending on the answers, different replication strategies are born so far[2,3,5]. In this paper, we propose a replication strategy for dynamic data grids that helps to increase file availability, to improve the response time by identifying

 https://sites.google.com/site/ijcsis/
ISSN 1947-5500

popular files in the specified time intervals and replicating these files on the appropriate sites.
The rest of the paper is organized as follows: the second part deals with the review of literature and conducted studies in this area. The third part outlines the proposed strategy. Simulation and its results will be discussed in section four and conclusion and some suggestions for future studies are presented in part five.

II. RELATED WORK

Some recent studies have discussed the problem of dynamic replication in data grids. Some of these works will be surveyed in this section.

Kavitha Ranganathan et al.[11] present various traditional replication and caching strategies. (1) No replication, (2) Best Client: a replica is created at the best client that has the largest number of requests for the file, (3) Cascading replication: once popularity exceeds the threshold for a file at a given time interval, a replica is created at next level which is on the path to the best client, (4) Plain caching: the client that requests the file stores a replica of the file locally, (5) Caching plus Cascading Replication: this combines Plain caching and Cascading replication strategy, and (6) Fast Spread: replicas of the file are created at each node along its path to the client. They measure access latency and bandwidth consumptions of each strategy with simulation tool and their simulation results show that Cascade method and Fast Spread method had the best performance among the six strategies in terms of bandwidth consumption and access latency.

In [12], Sang Min Park and Tai Hoon Kim proposed a two-level method called *Bandwidth Hierarchy based Replication (BHR)* that was inspired from internet hierarchy. In this algorithm, it is assumed that the sites that are located near each other are in the same network area so that the bandwidth among the sites within an area is more than the bandwidth among the sites between the areas. Therefore, If the required file is located in the same region, less time will be consumed fetching the file. The BHR strategy reduces data access time by maximizing network level locality and avoiding network congestion. The BHR strategy has good performance only when the capacity of the storage element is small.

In another paper [13] a new replication algorithm named Modified BHR was proposed. Modified BHR is an extension of BHR sterategy which tries to replicate files within a region and stores the replica in a site where the file has been accessed frequently based on the assumption that it may require in the future. This algorithm increases the data availability by replicating files within the region and also storing them in the site where the file has been accessed frequently. The mean job execution time and network usage are reduced further from BHR

Chang et al. [14] had a different view of data replication. They proposed a new data replication approach called *Fragmented Replicas*, which only replicates needed partial contents of a file locally to save storage space. Obviously, this strategy will face some challenges. One of the challenges is updating fragmented replications that must be solved.

In [15] Salah et al. proposed an algorithm called 4PDRA, which is based on Temporal and Geographical locality. Their sterategy includes four phase so that 4PDRA in first phase deals with identifying the popular data. Calculating a suitable number of new replicas is done in second phase. The algorithm continues by placing replicas and replacing old replicas with new replicas in third and fourth phase. The matrices used for evaluation of performance of 4PDRA are Mean Job Execution time (MJET), Average Storage Used and Effective Network Usage. The simulation results indicate that 4PDRA has better performance in comparison with No Replication, LRU, LFU, in terms of job execution time, effective network usage and percentage of storage filled.

III. DYNAMIC REPLICATION OF POPULAR FILES

In this section, we propose a novel dynamic replication strategy, called DRPF. As identified in Fig.1, the architecture of proposed algorithm has a three-level structure. Grid Sites(GS) are at the lowest level of the structure so that through the juxtaposition of sites that have similar interests and do similar tasks, the virtual organizations(VO) are formed. There is a Local Server(LS) within each VO. LSs are connected via internet which has low bandwidth. Therefore, speed of data access within VO is larger than across VOs. Region is the highest level and each Region consists of one or more VOs. There is a regional server(RS) in each region that controls one or several virtual organizations. Regional servers are connected to each other through the internet. Therefore, the bandwidth between the regions is less than the bandwidth between virtual organizations and consequently the transferring files between them takes a long time.

In a data grid, the user's job requires access to large number of files. So, to maximize the data locality, dynamic replication is required. Our proposed algorithm tries to reduce the data access time and to improve the execution time of tasks in grid by identifying popular files in the specified time intervals and replicating these files on the appropriate sites. According to what was said, DRPF algorithm is executed as follows:

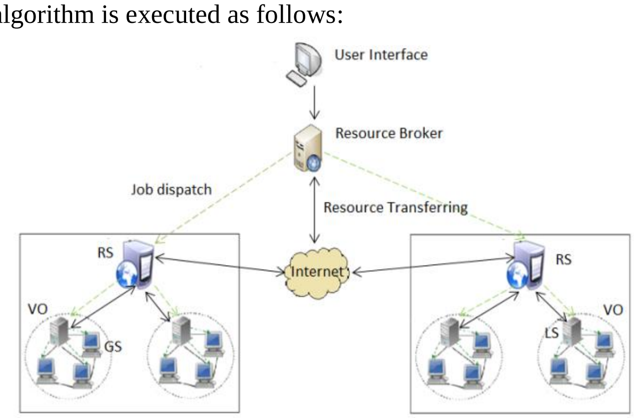


Figure 1. Architecture of the Proposed Method.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Since replication is a costly method, so first part of proposed algorithm is to identify the popular files in data grid. For this purpose, average number of requests for file f_j is calculated using the following equation:

$$Avg\text{-}num\text{-}of\text{-}\ request = \frac{\sum_{i=1}^{n} \#file-request}{n} \qquad (1)$$

Where n is the number of sites in VO. If the average number of requests is more than or equal to a predefined threshold, it is considered as a popular file. Now among these popular files, the ones are selected for replication that the number of their existing replicas is less than the maximum permitted number of replicas. The maximum value can be calculated using (2).

$$Max\text{-}rep = \frac{total\_space}{total\_files\_weight} \qquad (2)$$

Where total_space is the sum of all nodes' capacity and total_files_weight is the total size of all files in the data grid environment.

By identification of candidate files for replication, the algorithm goes to second step. Determination of the best sites for hosting new replicas is done in this part. In order to succeed this choice, our strategy takes into account the following parameters:

- Number of requests of each non-replication site for the desired file
- Bandwidth between replication sites and the non-replication sites

It should be explained, replication site is the site that is hosting the replica of the desired file, accordingly they never demand the files that are locally available; therefore, the new host sites should be far from replication sites as much as possible so that the workload can be distributed.

The competency of each site by using the following equation is obtained.

$$Competency\ s_i = num\_req + Avg\_Bandwidth \qquad (3)$$

$$Avg\_Bandwidth = \frac{\sum_{j=1}^{m-1} Bandwidth_{\ si,sj}}{m-1} \qquad (4)$$

Where s_i is the desired site for replication, s_j is replication site and m is the number of replication site. Also the criteria for calculating the average of competence all gridsites is defined by (5).

$$Avg\_Comp = \frac{\sum_{i=1}^{n} num\_req + Avg\_Bandwidth}{n} \qquad (5)$$

A site s_i is selected to replicate a file f_j if amount of it's compency is equal or greater than of average competence of total grid sites. Thus, the algorithm specifies a set of host site for each file f_j.

After determining the new host sites, new replicas of candidate files will be stored on these sites. Our sterategy first checks replica feasibility, in the other words proposed sterategy checks whether total size of the site's Storage Element is greater than or equal to the size of requested file. If it is not feasible, the file will be accessed remotely otherwise the requested file will be replicated. In this case if available storage size of applicant grid site is larger than the size of requested file, then the file can be replicated to applicant site, otherwise some of the existing replicas should be removed in order to store the new ones. As the storage capacities of grid sites are limited and also as replication itself is costly, data replication should be done carefully.

For this purpose, determining the value of replicas is a key factor that is defined differently in various algorithms. our sterategy via determine the value of replicas, tries to replace less valuable replicas with the new replicas. For this purpose, the value of replicas is determined based on the following four factors and is determined via (8):

i. The number of available replicas of the files in the virtual organization (NOR)

ii. The cost of doing replication (TC): it is attempted to keep the large files.

iii. The number of accesses to the replica in the past (NRR)

vi. The last access time to the replica(LAT):

LAT: $T_{\text{current}} - T_{\text{last access time}}$

$$TC = \frac{file\ size}{Bandwidth\ si,sj} \qquad (6)$$

Where BandWidth $_{si,sj}$ is available bandwidth between grid site 'i' (source grid site) and grid site 'j'(destination grid site).

$$NRR = \frac{\#\ request\ fj}{Tcurrent - Tstor} \qquad (7)$$

Where #request $_{fj}$ is the number of request for file f_j.

$$Replica\ Value = w_1 * TC + w_2 * \frac{1}{LAT} + w_3 * NRR + w_4 * \frac{1}{NOR} \qquad (8)$$

W coefficients are the rates of importance given to each factor by algorithm. The algorithm arranges the replicas in an ascending order based on their values and begins to delete from the top of the list until there is enough space to place the new replication in the target site.

In Fig. 2 DRPF algorithm is shown. DRPF tries to improve locality in accesses through increasing the the number of replicas in the virtual organization. The access frequency gathered by local server in VO and whitin each Grid Site because by maintaining and integrating them, the algorithm uses the more comprehensive and accurate information for replicate files.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

DRPF Algorithm:

Keep track of name of files and access frequently of each file whitin VO

by Local Server and whitin each GridSite

1. **Specifying the popular files for replication**
   Popular Files: Set of the files that must be replicated.
   Popular Files= ∅
   for each file f_j in VO
   {
   Calculate the *Avg-num-of- request*
   If the *Avg-num-of- request* ≥ predefined threshold Then
   {
   Specify the number of existing replicas of fj
   If the number of existing replicas of fj ≤ max-rep Then
   Popular Files= Popular Files ∪ f_j
   }// end if
   }// end of for
2. **Specifying the best sites for hosting the new replicas**
   For each file f_j in Popular Files
   {
   Host sites: set of best sites for hosting new replicas of f_j
   Host sites= ∅
   For each site s_i in VO
   {
   Calculate the competency s_i
   If competency s_i ≥ Avg_competency Then
   Host site= Host site ∪ s_i
   }//end of second for
   Return Host site for f_j
   }// end of first for
3. **Placement new replica**
   If (f_j.size >= se . total storage size in s_i) then
   access file remotely
   Else if (f_j.size <= se . available storage siza in s_i)
   store new replica of file f in s_i
   Else
   { for each replica in s_i
   {calculate $Replica\ Value = w_1 * TC + w_2 * \frac{1}{LAT} + w_3 * NRR + w_4 * \frac{1}{NOR}$}
   Selected replicas= sort all of replicas in s_i in ascending order,
   according to their " replica values"
   While (Selected replicas don't empty)
   { Select a replica from top of the " selected replicas " and remove it
   from s_i
   If (se . available storage size s_i >= f_j.size)
   { store new replica of file f in s_i ; exit ;}
   }// end of while
   }//end of else

Figure 2. DRPF Algorithm.

IV. SIMULATION AND EVALUATION OF RESULTS

In order to simulate and evaluate the efficiency of the proposed algorithm, OptorSim simulator has been used [16]. This simulating package which is written in Java is developed as a part of European Data Grid (EDG) project and in order to examine the efficiency of different replication algorithms in data grids.

As shown in Fig. 3 [17] OptorSim has several parts:

- Computing Element (CE): represents computational resource to which jobs can be sent in Data Grid.
- Storage Element (SE): represents data resource where data can be kept in Data Grid.
- Resource Broker (RB): schedules jobs to CEs according to scheduling algorithm
- Replica Manager (RM): at each site controls data transferring.
- Replica Optimiser (RO): within which a Replica Optimiser contains the replication algorithm which drives automatic creation and deletion of replicas.

Each site may provide computational and data resources. Ces run jobs by processing data files, which are stored in the SEs.

The grid configuration that we have used in our simulation is the CMS Data Challenge 2002 test bed [16] Fig. 4. For the CMS test bed, CERN and FNAL were given SEs of 100 GB and no CEs. All master files were stored at one of these sites.The simulated grid used in our experiments has 20 sites, 18 of them have Storage Element (SE) and Computing Element (CE) and 2 of them have only SE. The storage capacity of the master site is 200 GB and the storage capacity of all other sites is 50 GB. There are 6 job types, and each job type on average requires 15 files for execution. The size of single file is 1 GB.Therefore, total size of data in this configuration is 97 GB. The general parameters of our simulation are shown in Table I. We ran 200 jobs with six different job types.The simulation is repeated for 20 times and the final results are averaged.

The algorithm has been evaluated by comparing it with Least Recently Used(LRU), Least Frequently Used (LFU) and 4PDRA. In order to examine the effect of different access patterns in each replication stage each algorithm was performed with three different access patterns. Each job has a set of files it may request. The order in which those files are requested is determined by the access pattern. The following access patterns were considered in simulation:

Sequential: the set is ordered, forming a list of successive requests.

Gaussian random walk: files are selected from a Gaussian distribution centered on the previous file.

Random Zipf: files are selected using a Zipf-like distribution.

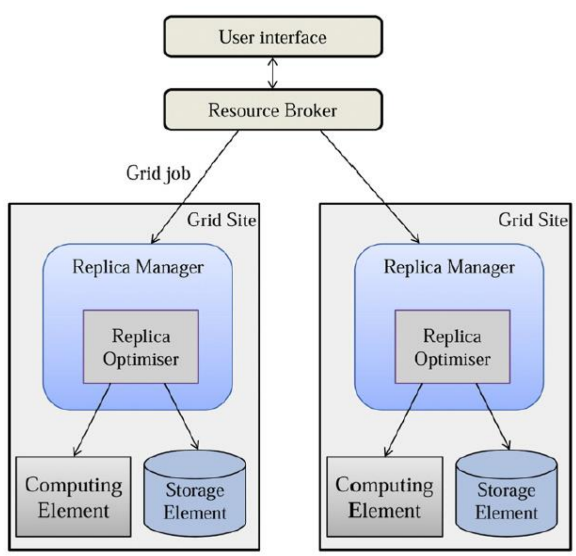


Figure 3. Optorsim Architecture[16].

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

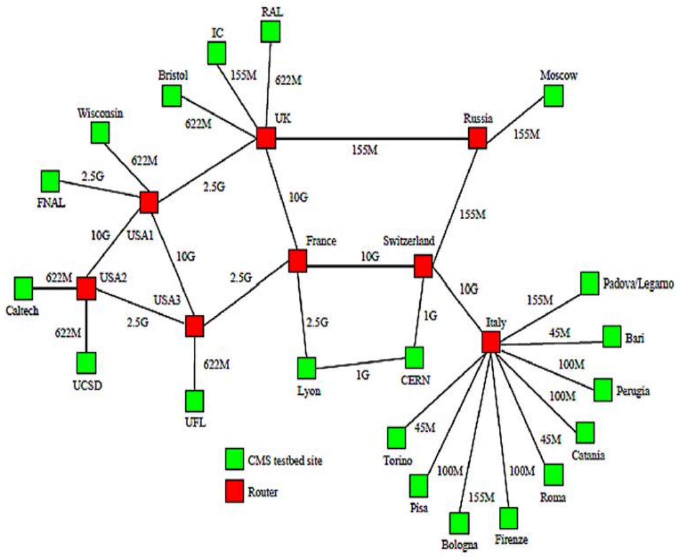


Figure 4. CMS Data Challenge 2002 Grid topology.

Table I. Simulation parameters.

Parameter	value
Number of Jobs types	6
Number of jobs	200
Number of file access per jobs	15
Job delay (ms)	2500
Number of sites	20
Number of storage elements (SEs)	20
Number of computing elements (CEs)	18
Access history length (ms)	10^6
Size of single file (GB)	1
Total size of files (GB)	97
Minimum bandwidth between sites (Mbit/s)	45
Maximum bandwidth between sites (Mbit/s)	10000
Storage capacity at each site (GB)	50, 200

In order to evaluate the effectiveness of the different replication strategies implemented in OptorSim, we used the following metrics:

• Mean job execution time;
• Effective Network Usage;

Mean job execution time: Among the factors, the mean job time is more important than the other ones because as this amount is lower the algorithm is better and has done jobs in less time. The total job time consists of the time of data transferring and job execution. This factor is obtained by dividing the total runtime of all tasks (in millisecond) over the number of tasks. Fig. 5 shows the mean job time of the four replication strategies with three different access patterns. The results of the simulation show that DRPF has the lowest value of Mean Job Execution Time. The reason is that in this strategy, future needs of grid sites are pre-sended for them by identifying popular files and storing them in appropriate grid sites; therefore more numbers of files are stored locally at the time of need. Another reason is that proposed algorithm tries to keep more valuable replicas during the replacement because one of the important factors that decreases the grid site's job execution time is having their required files locally stored on their storage element. The DRPF algorithm by considering number of available replicas of the files, replication cost, number of accesses in the past and last access time, in replacement algorithm, made our method better than the others because it does not delete valuable files which results in preserving the valuable replicas.

Effective Network Usage: ENU is a criterion that evaluates the ratio of replicated files to local accesses. This value ranges from zero to one. A lower value represents that the network bandwidth is used more efficiently so that the lower value of ENU indicates that algorithm has had better performance. It can be measured by using (9):

$$ENU = \frac{N\,remote\,file\,access + N\,file\,replication}{N\,remote\,file\,access + Nloca\,fileaccess} \qquad (9)$$

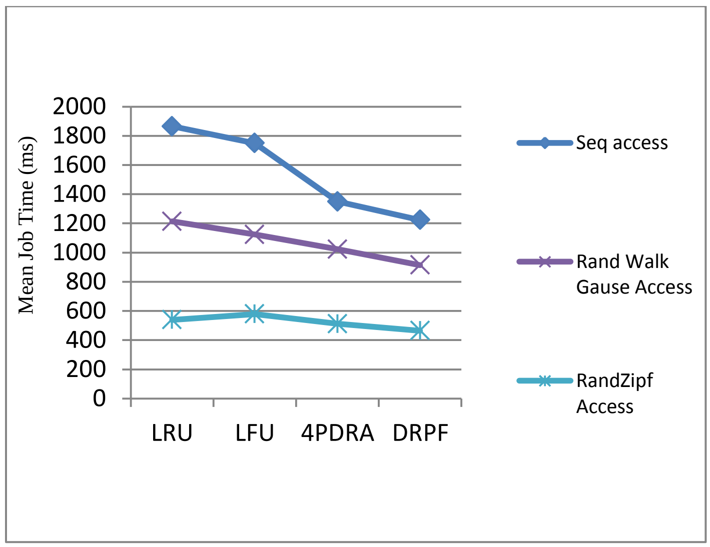


Figure 5. Mean job execution tim.

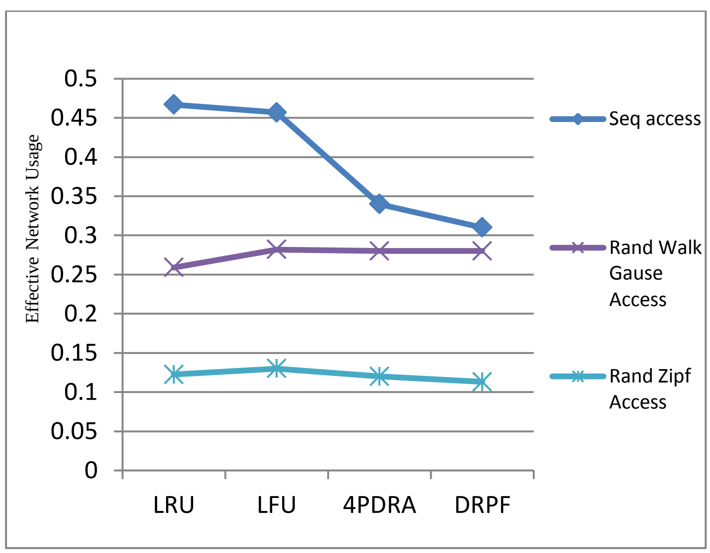


Figure 6. Effective network usage.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Fig. 6 shows the comparison of the Effective Network Usage of the four replication strategies. As it is obvious in this figure, DRPF has the lowest value of ENU in comparison with other methods. The reason is that LFU and LRU always replicate, so the large value of *N file replication* will increase the ENU value,while our sterategy by taking into account the popularity of files and the maximum number of replication for files, manages *N file replication* and also by increasing the the number of replicas in the virtual organization and preserving valuable replicas, improve locality in accesses. Therefore total number of replications and remote accesses has been decreased.

V. CONCLUSION AND FUTURE WORK

Data replication is a frequently used technique that can reduce bandwidth consumption and access latency in high performance data grids where end users demand remote accesses to large files. Since a grid environment is dynamic, network latency and user behavior may change.In this paper a new dynamic algorithm named DRPF for data replication in data grids was proposed. As grid sites within a VO have similar interest of files,our sterategy tries to replicate popular files as many as possible within a VO, where broad bandwidth is provided between sites. Therefore, sites will have their required files locally at the time of need and this will decrease response time, bandwidth consumption and increase system performance considerably. The proposed algorithm also tries to preserve the valuable replicas so that least valuable replicas should be replaced with new ones. The evaluation is based on four parameters including storage cost, number of available replications of the file in virtual organization, number of accesses to the file in the past and the last access time to the file. The results of simulation with OptroSim simulator indicate the efficiency of this method in data grid environment. The experimental results show that DRPF improves Mean Job Time and Effective Network Usage. As a future work, we plan to use the evolutionary algorithms to find the right place for replication. Also, We aim to predict the future needs of grid sites more accurately by taking advantage of data mining methods.

REFERENCES

[1] R.S. Chang and H.P. Chang, "A dynamic data replication strategy using access-weights in data grids," *Journal of Supercomputing*, Vol.45, Issue 3, pp. 277 – 295, 2008.

[2] Wuqing, Z. and Xianbin, X. and Zhuowei, W. and Yuping, Z. and Shuibing, H, "A Dynamic Optimal Replication Strategy in Data Grid Environment," *IEEE. International Conference on Internet Technology and Applications*,2010.

[3] Sang-Min Park, Jai-Hoon Kim, Young-Bae Ko, Won-Sik Yoon., "Dynamic Data Replication Strategy Based on Internet Hierarchy BHR," in*: Lecture notes in Computer Science Publisher, Springer-Verlag, Heidelberg*, *vol. 3033*,pp. 838–846, 2004.

[4] H. Lamhamedi, B. Szymanski, Z. Shentu, E. Deelman, "Data replication strategies in grid environments, *"In Proceedings of the 5th nternational Conference on Algorithms and Architectures for Parallel Processing* (ICA3PP'02). IEEE CS Press, Los Alamitos, CA, USA.

[5] A.R. Abdurrab, T. Xie, "FIRE: a file reunion data replication strategy for data grids," in: *10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing*, 2010, pp. 215–223.

[6] E. Deelman, H. Lamehamedi, B. Szymanski, S. Zujun,"Data replication strategies in grid environments," in: *Proceedings of 5th International Conference on Algorithms and Architecture for Parallel Processing,ICA3PP'2002, IEEE Computer Science Press, Bejing, China*, 2002, pp. 378–383.

[7] Sepahvand R., Horri A., Dastghaibyfard Gh, "A New 3-layer Replication and cheduling Strategy in Data Grid," *International Symposium on Telecommunications*, pp. 464-469, 2008.

[8] S.Y. Ko, R. Morales, and I. Gupta, "New worker-centric scheduling strategies for data-intensive grid applications," *Proc. ACM/IFIP/USENIX Int'l Conference on Middleware*, 2007, pp. 121-142.

[9] L. Meyer, J. Annis, M. Wilde, M. Mattoso, and I. Foster, "Planning spatial workflows to optimize grid performance," *Proc. ACM Symp.Applied Computing*, pp. 786 - 790, 2006.

[10] H. E.AL-Mistarihi and C. Yong, "Replica management in data grid," *International Journal of Computer Science and Network Security*, vol. 8, pp. 22–32, 2008.

[11] K. Ranganathan, I. Foster, "Identifying dynamic replication strategies for a highperformance data grid," in: *Proceedings of the Second International Workshopon Grid Computing, November 12*, pp. 75–86, 2001.

[12] Sang-Min Park, Jai-Hoon Kim, Young-Bae Ko, Won-Sik Yoon, "Dynamic Data Replication Strategy Based on Internet Hierarchy BHR,"*in: Lecture notes in Computer Science Publisher, vol. 3033, Springer-Verlag, Heidelberg*, 2004,pp. 838–846.

[13] K. Sashi, A.S. Thanamani, "Dynamic replication in a data grid using a modified BHR region based algorithm, " Future Generation Computer Systems 27, pp.202–210, 2011.

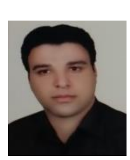
[14] R.S. Chang and P.H. Chen, "Complete and fragmented replica selection and retrieval in data grids," *Future Generation Computer Systems*, Vol. 23, Issue 4, pp. 536-546, May 2007.

[15] A. Salah, R. Javidan and M. Taghi, " A four-phase data replication algorithm for data grid," *Journal of Advanced Computer Science &Technology*, vol.4 ,pp. 163-174, 2015.

[16] W.H. Bell, D.G. Cameron, L. Capozza, .P.Millar, K. Stockinger, F. Zini,"Simulation of dynamic grid replication strategies in OptorSim*," International Journal of High performance Computing Applications* 17 (4) 2003.

[17] D.G. Cameron, A.P. Millar, C. Nicholson, "OptorSim: a simulation tool for scheduling and replica optimization in data grids," Proceedings of Computing in High Energy and Nuclear Physics (CHEP) (2004).

Soheila Malmoli Abbasi received her B.S. degree from Institute of Higher Education Khouzestan university jahad,Ahvaz, Iran in Computer Engineering in 2011. Since 2013 she is an M.S.student in Department of Computer Engineering, Khouzesan Science and Research Branch,Islamic Azad University. Ahvaz, Iran. Her research interests include grid computing and distributed systems.



Mohammadreza Noorimehr received the B.S. degree in Computer Engineering from Dezfool Azad University,Dezfool,Iran in 2006 and his M.S. Degree from the Araak Azad University degree in computer engineering,Arrak,Iran in 2009,respectively. He also received the Ph.D. Degree in computer architecture from Sience and Research Branch of IAU,Tehran,Iran in 2013. He is currently Assistant professor in faculty of computer engineeringof Ahvaz University. His research intrests include computer arithmetic,distributed systems and interconnection network.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Image Steganography Method for Concealing Secret Data into Coefficients Based on High Scalable Sub-Bands of Integer Wavelet Transform

YAHYA E. A. AL-SALHI[a], SONGFENG LU[b]
[a,b]School of computer science,
Huazhong University of Science and technology,
Wuhan 430074,PR china
[a]I201422150@hust.edu.cn
[b]lusongfeng@hust.edu.cn
[a] Corresponding author

***Abstract*—In information security, an image steganography technique uses one of the most popular transforms; either a spatial domain or the frequency domain to conceal the secret information. In this paper, an image steganography system using the spatial domain technique to conceal secret information in the frequency domain is proposed to conceal secret image information in another cover image. The Integer Wavelet Transform (IWT) used to obtain high scalable sub bands for each LL, LH, HL and HH of the cover image file. Then, the steganography approach is used to conceal the secret information in the wavelet coefficients for all sub bands. The results show high quality of stego image, and the stego image is analyzed for different attacks. It is found that the technique is robust, and it can withstand the attacks. The quality of the stego image is measured by Peak Signal to Noise Ratio (PSNR), Structural Similarity Index Metric (SSIM), and Universal Image Quality Index (UIQI). The quality of extracted secret image is measured by Signal to Noise Ratio (SNR) and Squared Pearson Correlation Coefficient (SPCC).**

1. Introduction

For many years, image steganography has been a huge challenge for researchers [1]. Steganography conceals secret information in other medias (such as image, audio, video and etc.) [2]; known as cover files. Cover file along with the concealed image information is known as stego file. The secret file can be text message, image or audio. The steganography is achieved in transform domain [3]. There are two types of steganography techniques: temporal domain and transform domain. In the temporal domain, the actual sample values are manipulated to conceal the secret image information. In transform domain, the cover file is converted to the different domains such as a frequency domain; to get the transformed coefficients. These coefficients are manipulated to conceal the secret image file. Then the inverse transformation is applied on the coefficients to get stego-image file. The temporal domain techniques are easily to attacks than transform domain techniques; because there are actual sample values are modified. The transforms that can be used are Fast Fourier Transform (FFT) [4], Discrete Cosine Transform (DCT) or Discrete Wavelet Transform (DWT) [5], [6]. This paper discusses how the successive steganography concept is used in conjunction with DWT coding, and Haar transform to achieve image steganography [7], [8] because wavelet transformation gives frequency content of a function $f(t)$as a function of time. The drawback of FFT is that Fourier Transform gives frequency information, but it does not provide information about timings. This is because the basic functions (sine and cosine) used by this transform are infinitely long. They pick up the different frequencies of $f(t)$ regardless of where they are located.

This paper is organized as follows: Section 2 discuss the preliminaries which including how DWT theory and image steganography methodology for increasing sub-bands, and concealing the sub-bands into cover image file. This is important in image steganography because bands can be hidden in the cover file. Extremely important information despite that they represents in only a tiny fraction of the image samples. The previous related research work in DWT and image steganography is discussed in the section 3. Section 4 introduces new scheme of DWT with a new functionality to get unlimited sub-bands of the cover image and shows how it coding efficiently encode a significance bands of wavelet coefficients by predicting the absence of significant information across scales. Section 5, discusses experimental results for different rates and for various standard test images. The paper concludes with the section 6.

2. Preliminaries

2.1. Integer Wavelet Transform (IWT)

In wavelet transformation, an integer wavelet transform is selected, a function that is Haar [8] in some interval, and it is used to explore the features of the function $f(t)$ in that interval. The IWT is converted to another interval of time and used in the same way. So with IWT, sub-bands be scaled

to provide a time-frequency representation of the sub-bands. There are many wavelets transform discovered [9]. The simplest one is the Haar wavelet transform [8]. Information that is produced and analyzed in real-life situations is integered. It can be in the integer form of numbers, rather than a Haar function. That is why the discrete rather than the interger wavelet transform is in practice. When the entired image data involve sequences of integers as in the case for images, wavelet transforms can use the map integers-to-integers. In steganography method, the significant objects to conceal the secret objects is an image it is call cover image or cover file [10]. The cover image can be in gray scale form (threshold level) or color form. Color images can be represented in various formats such as Red Green Blue (RGB), Hue Saturation Value (HSV); YUV, YIQ and YCbCr (Luminance, Chrominance). Color image steganography can be done in any color space domain. n transformation, there are two types of domains , first one is fourier domain and the ssecond one is frequency domain [11]. The frequency domain is used to conceal the secret data into the cover image file according to the IWT. When the IWT applied in the color images, the coefficients of transforms obtained for all the three channels in the corresponding representation [2]. Wavelet transform is applied to an image, decomposed into four sub-bands LL,LH,HL and HH. The low-frequency sub-band is LL and contains the approximation coefficients [8] The significant features of the image are stored in the approximation coefficients [2]. Other three sub-bands are high-frequency sub-bands and contain fewer significant features [5], [7]. It is possible to reconstruct the image by considering only LL sub-band. When secret image samples are transformed, approximation and detailed coefficients are produced. Approximation coefficients contain the most significant features. In this case, it is possible to reconstruct the secret image by considering only approximation coefficients.

2.2. Characteristics of Image Steganography

For images, it is measured in terms of Peak Signal to Noise Ratio (PSNR), Structural Similarity Index Metric (SSIM) [12], Universal Image Quality Index (UIQI), Color Image Quality Measure (CQM) etc. For sub-bands it is measured in terms of Signal to Noise Ratio (SNR) and Squared Pearson Correlation Coefficient (SPCC) etc. [13], [14].

2.2.1. Peak Signal to Noise Ratio (PSNR). The calculation of PSNR is given in “Eq. (1);

$$PSNR = 10.log_{10}\left(\frac{MAX^2}{MSE}\right), \quad (1)$$

where MAX is the maximum value of pixels (255 for grey scale images). MSE is the mean square error between the original and stego images. It is given in “Eq. (2);

$$MSE = \frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\|O(i,j) - S(i,j)\|^2, \quad (2)$$

where $O_{i,j}$ is original pixel and $S_{i,j}$ is stego pixel. Greater PSNR values indicate better quality. It is expressed in decibels (dB).

2.2.2. Structural Similarity Index Metric (SSIM). SSIM is an objective image quality metric and superior to traditional measures such as MSE and PSNR [15], [16]. PSNR estimates the perceived errors, whereas SSIM considers image degradation as perceived change in structural information. Structural information is the idea that the pixels have strong inter-dependencies, especially when SSIM are spatially close. These dependencies carry important information about the structure of the objects in the visual scene. SSIM is given in “Eq. (3);

$$SSIM = \frac{(2\bar{x}\bar{y} + c_1)(2\sigma_{xy} + c_2)}{(\sigma_{x^2} + \sigma_{y^2} + c_2)(\bar{x}^2 + \bar{y}^2 + c_1)} \quad (3)$$

where $C_1 = (K_1L)$, and $C_2 = (K_2L)$ are two constants used to avoid null denominator. L is the dynamic range of the pixel values (typically, these are $2^{bitsperpixel}$ bits per pixel -1). $K_1 = 0.01$ and $K_2 = 0.03$ by default. The dynamic range of $SSIM$ is between [-1 and 1]. Maximum value of 1 will be obtained for identical images.

2.2.3. Universal Image Quality Index (UIQI). UIQI is also an objective image quality measure. It is given in “Eq. (4);

$$UIQI = \frac{(4\bar{x}\bar{y}\sigma_{xy})}{(\sigma_{x^{n-1}} + \sigma_{y^{m-1}})(\bar{x}^{m-1} + \bar{y}^{n-1})} \quad (4)$$

where $\bar{x}$, $\bar{y}, \sigma_x, \sigma_y$ and σ_{xy} are given by Eqs: (5, (6), (7), (8) and (9) respectively;

$$\bar{x} = \frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N}(x(i,j)) \quad (5)$$

$$\bar{y} = \frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N}(y(i,j)) \quad (6)$$

$$\sigma_{x^{m-1}} = \frac{1}{M \times N - 1}\sum_{i=1}^{M}\sum_{j=1}^{N}(x(i,j)\bar{x}) \quad (7)$$

$$\sigma_{y^{\infty}} = \frac{1}{M \times N - 1}\sum_{i=1}^{M}\sum_{j=1}^{N}(y(i,j)\bar{y})^{\infty} \quad (8)$$

$$\sigma_{xy} = \frac{1}{M \times N - 1}\sum_{i=1}^{M}\sum_{j=1}^{N}((x(i,j) - \bar{x})(y(i,j)\bar{y})) \quad (9)$$

where M and N are maximum and minimum values of (*i,j*).
This quality index represents any distortion as an amalgamation of three factors: loss of correlation, luminance distortion and contrast distortion. To illustrate it, the definition of $UIQI$ can be written as a product of the three components:
see "Eq. (10):

$$UIQI = Q_1 \times Q_2 \times Q_3 \tag{10}$$

where Q_1, Q_2 and Q_3 are given in "Eqs. (11), (12) and (13) respectively;

$$Q_1 = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{11}$$

$$Q_2 = \frac{2\bar{x}\bar{y}}{\bar{x^{m-1}} + \bar{y^{n-1}}} \tag{12}$$

$$Q_3 = \frac{2\sigma_x \sigma_y}{\sigma_{\bar{x}^{m-1}} + \sigma_{\bar{y}^{n-1}}} \tag{13}$$

Q_1 represents the correlation coefficient between x and y, which is the measure of degree of linear correlation of x and y. Whereas Q_2 indicates luminance closeness between x and y. Q_3 denotes contrast similarities between the two images. The dynamic range of $UIQI$ is between [-1 and 1]. For identical images, its value will be 1.

2.2.4. Color Image Quality Measure (CQM). The CQM is given in "Eq. (14);

$$CIQM = PSNR_Y \times R_W + \left(\frac{PSNR_U + PSNR_V}{2}\right) \times C_W \tag{14}$$

where $PSNR_Y$, $PSNR_U$ and $PSNR_V$ are the $PSNR$ values of Y,U and V components of the color image respectively. R_W and C_W are the weights on the human perception of cone and rod sensors respectively. In HVS cones are responsible for chrominance perception and rods are responsible for luminance perception. $C_W = 0.0551$ and $R_W = 0.9449$ as specified by HVS. CQM greater value indicates greater image similarity. It is represented in dB.

2.2.5. Signal to Noise Ratio. The SNR is given in "Eq. (15);

$$SNR = 10.log_{10}\left(\frac{\frac{1}{M \times N}\sum_{i=1}^{M} x_i \sum_{j=1}^{N} y_i}{MSE}\right)^{m-1} \tag{15}$$

where MSE is given in "Eq. (16),

$$MSE = \frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N} \| \tag{16}$$

where $x_i - y_j$, x_i is the original sample and y_j is the stego sample.

SNR refers to the measurement of the level of an sub-bands as compared to the level of noise that is presents in that band. The measurement is usually expressed in decibels (dB). A larger value of SNR implies a better quality. However, it is a statically measured quantity and so does not judge the quality as a whole.

2.2.6. Squared Pearson Correlation Coefficient (SPCC). SPCC measures the similarity level between two bands [2], [17], [18]. The SPCC uses to measure the similarity level among the sub-bands using many filters [17], [18]. The higher the SPCC, the higher is the similarity level. Its range is between 0 and 1. It is given in "Eq. (17);

$$SPCC = \left[\frac{\sum_{i=1}^{n}(x_i - \ddot{x})(y_i - \ddot{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \ddot{x})^{m-1}}\sqrt{\sum_{i=1}^{n}(y_i - \ddot{y})^{n-1}}}\right]^{m-1} \tag{17}$$

where x_i and y_i are the two bands, $\ddot{x}$ and $\ddot{y}$ are their averages.

3. Related research works

Frequency and Fourier has been discussed [7], [19], to conceal the secret image in sub-bans of the cover image [5]. Human Visual System (HVS) is sensitive to perciptive changes in luminance but not in chrominance [20]. YCbCr is one of the representations where Y is the luminance, Cb and Cr are the chrominance components [9]. The chrominance can be modified, without any visually damaging the overall image quality [2]. From variouse researchs, it is found that the secret file is concealed in the cover file. The quality of performance analysis of an image steganography based on IWT associated to the color and gray scale images was proposed in (1993) by Shapiro, J.M [7]. Vijay Kumar and Dinesh Kumar [2], [21], proposed a steganography method intended to observe the effects of concealing the secret message in different sub-bands coefficients such as CH, CV and CD [5], on the performance of stego image in terms of Peak Signal to Noise Ratio ($PSNR$), with distortion tolerance. It uses the fourier domain for concealing the secret image. This scheme provides distortion tolerance and gives high quality of processed image [8]. Mean sequre error (MSE) and Peak Signal to Noise Ratio ($PSNR$) are acceptable for image similarity measure only when the images are variate by increasing the distortion of a certain kind. However, they fail to capture image quality when they are used to measure across distortion kinds. Structural Similarity Index Metric $SSIM$ and Universal Image Quality Index ($UIQI$) are widely used method for measurement of image quality based on Human Visual System (HVS) [13], [14]. The approximate sub-band coefficients are generated by certain integer wavelet transform to have image qualities [7], [13]. High capacity of the image steganography technique, depends on integer wavelet transform with an acceptable different levels of imperceptibility and distortion

is proposed [14]. Concealed block coding with the optimized truncation of the concealed bit-streams (EBCOT) proposed by David Taubman for Scaling an Image Compression with EBCOT [11]. The Steganography method based on Integer Wavelet Transform was proposed for concealing the secret image in sub-bans form [22].

4. Proposed scheme

In this section of research, image steganography method is proposed to conceal the secret image bands into the cover file. The secret and cover images can be in any format such as(jpg, bmp and etc). In the previous methods , only two bits of the secret message have been XORed with one byte of the cover file. Since, the secret image sub-bands have large number of samples even for small duration, the cover image has considerably large. Color images are suitable because of enough hiding space. Since YCbCr approach is more secure than RGB approach, while YCbCr approach is used [23]. The cover image is converted to YCbCr. Then Cb, Cr components and secret sub-bands are transformed using IWT. The approximate coefficients of the secret sub-bands are concealed in the second and third bit planes of high frequency coefficients of the Cb and Cr.

4.1. Concealing of n bits per coefficient of the Integer Wavelet Transform Domain

the Concealing Procedures are given in the following steps:
$\mathbf{Input}$: cover image CC_I and secret image S_I.
$\mathbf{Step1}$: Represent CC_I in gray-level and obtain IWT of sub bands of CLL, CHL, CLH and CHH. The cover file will be partitioned into scales of sub-bands. It is given in Eq. (18):

$$S(k_1,k_2)=\frac{1}{\sqrt{N_1N_2}}\sum_{k_1}^{n-1}\sum_{k_2}^{m-1}w_{\varphi}^{m-1}+WS(Hs) \quad (18)$$

where WS, HS, w_{φ}^{m-1} and w_{ψ}, are given in "Eqs. (19) , (20), (21) and (22)respectively;

$$WS=\sum_{k_1}\sum_{k_2}w_{\psi}^{i}(j_0k_1k_2)\psi_{j_0k_1k_2}(n_1,n_2) \quad (19)$$

$$HS=\frac{1}{\sqrt{N_1N_2}}\sum_{i=H,V,D}\sum_{j=0}^{m-1} \quad (20)$$

$$w_{\varphi}^{n-1}(j_0k_1k_2)=\frac{1}{\sqrt{N_1N_2}}\sum_{n_1=0}^{m-1}\sum_{n_2=0}^{n-1}S(n_1,n_2)\varphi_{j_0k_1k_2}(n_1,n_2) \quad (21)$$

$$w_{\psi}^{i}(j_0k_1k_2)=\frac{1}{\sqrt{N_1N_2}}\sum_{n_1=0}^{m-1}\sum_{n_2=0}^{n-1}S(n_1,n_2)\psi_{j_0k_1k_2}(n_1,n_2) \quad (22)$$

where
$i = H,V,D$ and S is a size of image and is equal to $N_1 \times N_2$
$\mathbf{Step2}$: Obtain IWT of secret image bands to get approximation and detail coefficients.
$\mathbf{Step3}$: Conceal the approximation coefficients of secret image bands in the second and third LSB planes of CHH and CLH sub bands after encryption.
In this situation, unlimited sub-bands of the secret message are concealed in bytes of the cover file. Suppose S_i and S_j are two secret bits, see "Eqs. (23) and (24);

$$S_i = S_i \quad XOR \quad b_{i-1} \quad XOR \quad b_{i-2}, i \leq m-1 \quad (23)$$

$$S_j = S_j \quad XOR \quad b_{j-1} \quad XOR \quad b_{j-2}, j \leq n-1 \quad (24)$$

where b_i and b_j are 4^{th} and 5^{th} bits of the cover byte respectively. The 2^nd and 3^rd bits of the cover byte are replaced by these encrypted secret bits. This type of dynamic encryption avoids the need for encryption key. Concealing can be done in the Cr component also in the similar fashion. Here C_1 and C_2 are the modified CLH and CHH. For concealing any two bits of the secret image in one byte of the cover file, two bits from the secret image are $XORed$ with two bits of the cover file. In our situation, the sub-bands of the secret file are $XORed$ with bytes of the cover file, it is given in Eqs. (25) and (26);

$$S_{m-1_i} = S_i \quad XOR \quad b_{i-1} \quad XOR \quad b_{i-2} \quad (25)$$

and

$$S_{n-1_j} = S_j \quad XOR \quad b_{j-1} \quad XOR \quad b_{j-2} \quad (26)$$

$\mathbf{Step4}$: Obtain inverse IWT to get stego Cb. Then convert to RGB system.
$\mathbf{Output}$: stego image G.

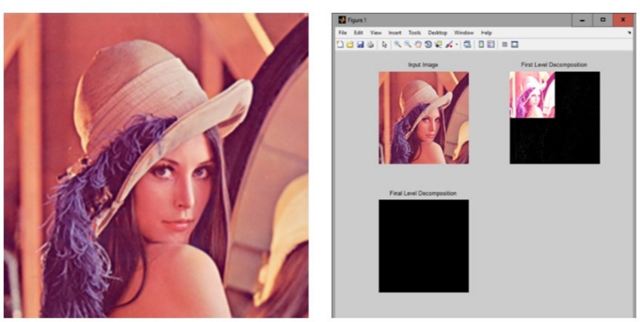
4.2. Extraction of n bits per coefficient of the Integer Wavelet Transform Domain

the Extracting Procedures are given in the following steps
$\mathbf{Input}$: Stego image G.
$\mathbf{Step1}$: Read stego image G and represent in the gray-level format.
$\mathbf{Step2}$: Obtain IWT of scalable sub bands:GLL,GHL,GLH, and GHH.
$\mathbf{Step3}$: Extract the encrypted secret sub-bands bits from the second and third bit planes of GLH and GHH.
In this method, unlimited sub-bands of the secret message are obtained from one byte of the stego image coefficient. Then decryption is done as follows: the two encrypted bits are $XORed$ with bits of the stego byte to get secret bits, i.e., as mentioned in "Eqs. (19), (20) and (21).
$\mathbf{Step4}$: Convert to decimal to get approximation

coefficients of secret sub-bands.
Step5 : Obtain inverse IWT for approximation coefficients obtained in Step 4 and considering zeroes for detailed coefficients. The result is secret image bands.
Step6 : End extracting.
Output : Secret image file **S**.

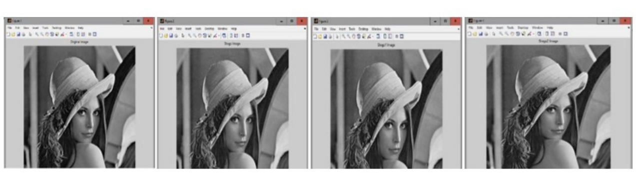
5. Experimental Results and Performance Evaluation

In this paper, the algorithm is tested by using 512×512 size of color image. It is transformed by using Integaer wavelet transform (IWT) to obtain last level of the sub-bands for secret image to be concealed into an image file (cover file). When the payload capacity is decreased to 131128 and 262256 sub-bands, only two levels of integer wavelet transformation are performed, so that the approximate coefficients of sub-bands to be concealed are reduced to one eighth. Many levels of inverse wavelet transformation sub-bands in extracting process are performed. Here, different image formats can be used such as (jpeg, png, bmp and etc.). There is no effect of the images format on the performance evaluation metrics because both cover and stego images can be in any format and data concealing is done in the transform domain.We will present the experimental results of our proposed method, which has been implemented in **MATLABR2011a**. Fig. 1 shows the original image and cover file after transformed using the unlimited sample of sub-bands. Figs. 2 and 3shows the plots of the stego image and extracted secret bands respectively.



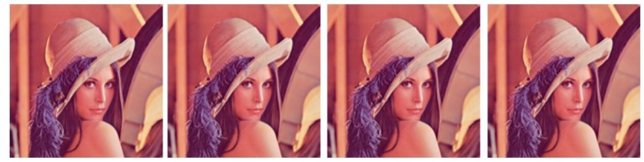
(a) (b)

Figure 1. (a) original image (b) cover file with the high scalable sub-bands of IWT transform.



(a) (b) (c) (d)

Figure 2. Original and Stego images: (a)Original, (b)Stego with 131128 samples, (c) Stego with 262256 samples, (d) Stego with 524512 samples.



(a) (b) (c) (d)

Figure 3. Cover and Extracted images: (a) Cover, (b) Extracted from 131128 samples, (c) Extracted from 262256 samples, (d) Extracted from 524512 samples

In Fig. 3, the comparisons on hiding payloads are presented by the size of the concealing data (in bits) vs. the image distortion, as presented by the peak signal-to-noise ratio (PSNR) (in dB) see “Eq. (27).

The performance evaluation metrics for the stego images and extracted secret images sub bands are shown in Table1. The stego sub-bands are evaluated using $PSNR$, $SSIM$ and CQM. $UIQI$ and $SSIM$ values obtained are same, therefore it is not included in the Table 1.
The extracted secret image bands are evaluated using SNR and $SPCC$. It is observed that when the secret sub-band samples are increased above 262256 , the quality of the stego and the extracted secret histograms are decreased below the HVS and HAS limits. It is because two levels of wavelet transformation are taken before concealing the secret message.
In this case, the extracted secret bands differed slightly. Otherwise, it is exactly same as the original.

TABLE 1. PERFORMANCE METRICS FOR THE STEGO AND EXTRACTED SECRET IMAGE SUB-BANDS OF SAMPLES

Cover image	Secret		Stego		Extracted	
512×512	samples	PSNR	SSIM	CQM	SNR	SPCC
lena.jpg	524512	66.30	13.30	55.68	38.3	0.9022
lena.jpg	262256	38.6	12.34	52.68	36.3	0.8922
lena.jpg	131128	38.7	12.34	52.68	32.4	0.8353

5.1. Results discussion and evaluation

5.1.1. Concealing Payload Evaluation. Concealing payload is a basic measurement to evaluate the steganography scheme performance. Payload refers to the amount of bits that can be concealed into the cover image. High concealing payload is computed using “Eq. (27)

$$E_p = \frac{Max}{C_{I_{WH}}}(bit\ per\ pixel) \tag{27}$$

where Max is the maximum number of the secret message bits that can be concealed into cover image, and $C_{(I_{W,H})}$ are the cover image width and height respectively. Where, all of them are measured by bit per pixel (BPP).

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

5.1.2. Concealing distortion Evaluation. Concealing distortion used to evaluate the stego image quality using Peak Signal-to-Noise Ratio PSNR which is calculated using “Eq. (28).

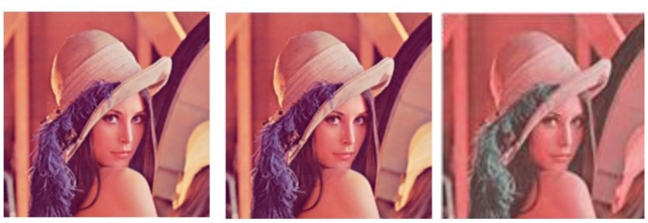
$$PSNR = 10log_{10}[\frac{255^2}{MSE}](dB) \tag{28}$$

Where MSE is the Mean Square Error, and is refers to the difference value between cover and stego images, which is given in “Eq. (29):

$$MSE = \frac{1}{C_{I_{WH}}} \sum_{i=1}^{W} \sum_{j=1}^{H} (C_{ij} - S_{ij})^2 \tag{29}$$

where $C_{(i,j)}$ and $S_{(i,j)}$ are the gray values of pixel (i, j) of the cover and stego images. W and H are the width and height of the cover image (the stego image has the same size).

By considering one cover image and varying the secret sub-band samples, the results can be analyzed easily. In this technique, the maximum payload size 524512 samples (each sample is 8 bits), with 512×512 color image. If both Cr and Cb components are used to hide the secret messages, then the quality of the stego decreases and secret message cannot be hidden to the maximum extent possible without crossing the quality metrics limits. This work compared, where $SSIM$ is measured.



(a) (b) (c)

Figure 4. Effect of attacks: (a) NO attack, (b) Gaussian, (c) Median filtering.

TABLE 2. SHOWS THE SNR AND SPCC OF THE EXTRACTED SUB-BANDS

Attack type	SNR in dB	SPCC
No attack	38.3	0.9022
Gaussian noise	37	0.9022
Median filtering	36	0.9020

5.1.3. Analysis for common attacks. Testing the performance of our scheme is necessary when the algorithm was designed, by subjecting it to different types of attacks. It should be possible to retrieve the concealed image data even if the stego image undergoes certain attacks. The most popular attacks that the stego-image may experience is Gaussian noise, median filtering, JPEG compression, scaling, cropping, etc. JPEG compression and image scaling may not affect the stego-image and extraction process, because concealing is done in fourier domain of an integer wavelet transform. Here two most popular attacks are considered: Gaussian noise and median filtering. Gaussian noise attack is performed with zero mean and 0.001 variance. Median filtering is performed using 3-by-3 neighborhood. In both cases, the secret sub-bands can be obtained with reasonable SNR and SPCC values. The stego images before and after the attacks are shown in Fig.4 and Table 2.

6. Conclusion

In this paper, a secure robust and the final scale bands are generated of the secret image is proposed. It gives good values for all the metrics, and hence this is an enough method to send band samples without revealing its existence. The performance of the proposed scheme against some of the attacks is also good. The scheme needs to be tested to prevent other attacks like histogram equalization, cropping, occlusion, translation and etc. The experimental results show that the secret sub-bands can be extracted without much distortion in most of the cases.

Acknowledgments

The authors gratefully acknowledge the support from the National Natural Science Foundation of China under (Grant No. 61173050).

References

[1] C. L. P. Chen, M.-C. Chen, S. Agaian, Y. Zhou, A. Roy, and B. M. Rodriguez, A pattern recognition system for JPEG steganography detection, Optics Communications, vol. 285, no. 2122, pp. 4252-4261, 10/1/, 2012.

[2] A. Ioannidou, S. T. Halkidis, and G. Stephanides, A novel technique for image steganography based on a high payload method and edge detection, Expert Systems with Applications, vol. 39, no. 14, pp. 11517-11524, 10/15/, 2012.

[3] Ahmed, N.; Natarajan, T.; Rao, K.R., ”Discrete Cosine Transform,” in Computers, IEEE Transactions on , vol.C-23, no.1, pp.90-93, Jan. 1974.

[4] Ali Al-Ataby, Fawzi Al-Naima. A Modified High Capacity Image Steganography Technique Based on Wavelet Transform. The International Arab Journal of Information Technology, Vol. 7, No. 4, October 2010.

[5] Shapiro, J.M., ”Embedded image coding using zerotrees of wavelet coefficients,” in Signal Processing, IEEE Transactions on , vol.41, no.12, pp.3445-3462, Dec 1993.

[6] Glassman, J.A., ”A Generalization of the Fast Fourier Transform,” in Computers, IEEE Transactions on , vol.C-19, no.2, pp.105-116, Feb. 1970.

[7] Jeremiah Spaulding, Hideki Noda, Mahdad N. Shirazi, Michiharu Niimi, Eiji Kawaguchi. BPCS Steganography Using EZW Encoded Images. DICTA2002: Digital Image Computing Techniques and Applications, 2122 January 2002, Melbourne, Australia.

[8] M. Chaumont, W. Puech, C. Lahanier, Securing color information of an image by concealing the color palette, Journal of Systems and Software, Volume 86, Issue 3, March 2013.

[9] S. Hemalatha, U. D. Acharya, and A. Renuka, Wavelet Transform Based Steganography Technique to Hide Audio Signals in Image, Procedia Computer Science, vol. 47, pp. 272-281, //, 2015.

[10] Shikha Sharda, Sumit Budhiraja. Performance Analysis of Image Steganography based on DWT and Arnold Transform. International Journal of Computer Applications (0975 8887)Volume 69 No.21, May 2013.

[11] Suvarna Patil, Gajendra Singh Chandel, Ravindra Gupta. Performance Analysis of Steganography based on 5-Wavelet Families by 4 Levels DWT. IJCSNS International Journal of Computer Science and Network Security, VOL.14 No.12, December 2014.

[12] Rezazadeh, S.; Coulombe, S., ”A novel approach for computing and pooling Structural SIMilarity index in the discrete wavelet domain,” in Image Processing (ICIP), 2009 16th IEEE International Conference on , vol., no., pp.2209-2212, 7-10 Nov. 2009.

[13] H. Al-Dmour, and A. Al-Ani, A steganography embedding method based on edge identification and XOR coding, Expert Systems with Applications, vol. 46, pp. 293-306, 3/15/, 2016.

[14] S. Hemalatha, U. Dinesh Acharya, A. Renuka, Wavelet Transform Based Steganography Technique to Hide Audio Signals in Image, Procedia Computer Science, Volume 47, 2015.

[15] Jian-Jiun Ding; Soo-Chang Pei; Po-Hung Wu, ”Jacket Haar transform,” in Circuits and Systems (ISCAS), 2011 IEEE International Symposium on , vol., no., pp.1520-1523, 15-18 May 2011.

[16] Jiaolong Yu; Benesty, J.; Gongping Huang; Jingdong Chen, ”Examples of optimal noise reduction filters derived from the squared Pearson correlation coefficient,” in Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on , vol., no., pp.1552-1556, 4-9 May 2014.

[17] Mandal, D.; Panetta, K.; Agaian, S., ”Human visual system inspired object detection and recognition,” in Technologies for Practical Robot Applications (TePRA), 2012 IEEE International Conference on , vol., no., pp.145-150, 23-24 April 2012.

[18] Jianqin Zhou; Ping Chen, ”Generalized Discrete Cosine Transform,” in Circuits, Communications and Systems, 2009. PACCS ’09. Pacific-Asia Conference on , vol., no., pp.449-452, 16-17 May 2009.

[19] Fouroozesh, Zohreh. ”Image steganography based on LSBMR using Sobel edge detection.” In e-Technologies and Networks for Development (ICeND), 2014 Third International Conference on, pp. 141-145. IEEE, 2014.

[20] Taubman, D., ”High performance scalable image compression with EBCOT,” in Image Processing, IEEE Transactions on , vol.9, no.7, pp.1158-1170, Jul 2000.

[21] M. Iranpour, ”A novel steganographic method based on edge detection and adaptive multiple bits substitution.” pp. 1-6.

[22] Kumar, V.; Kumar, D., ”Performance evaluation of DWT based image steganography,” in Advance Computing Conference (IACC), 2010 IEEE 2nd International, vol., no., pp.223-228, 19-20 Feb. 2010.

[23] N. N. El-Emam, and M. Al-Diabat, A novel algorithm for colour image steganography using a new intelligent technique based on three phases, Applied Soft Computing, vol. 37, pp. 830-846, 12//, 2015.

Managing and Tracking Alumni in Saudi Universities

Dr. Amr Jadi
Department of Computer Science and Software Engineering
College of Computer Science and Engineering, University of Hail
Hail, Saudi Arabia

***Abstract*— Managing Alumni System is one of the greatest challenges in the present market of Saudi Arabia. An alumni system is a channel between different universities and labor market to deliver various services to students as per the merit and priorities. There is no constructive method in present system of Labor office to monitor job requests from the students and communicate them with potential changes of market policies. This research aims to provide an architecture building a Functional Alumni System in Saudi Universities. The loop holes of current alumni system are highlighted and a consolidated methodology is implemented to develop a unique approach for increasing challenges.**

To overcome these deficiencies between Alumni Systems and Labor Market, the preset research provides a runtime monitoring system based on Labor policies to attain quality and manageability. The requests placed by students, applications executed by labor office and job requests in pending can be monitored and processed with a flexible approach by using this method. In turn lot of financial wastage can be avoided by reducing the complexity between job seekers and providers by the proposed approach.

Keywords *- Runtime Monitoring, Policy, Alumni System, Saudi Universities, Labor Office, Integration*

I. Introduction

Universities in Saudi Arabia need to establish an Electronic Collaboration System (ECS) to suppress the challenges and major problems by integrating all government and private universities. Each student from different study area must be able to enroll and apply for the jobs available in the market. No matter from where the student belongs to, the ECS must be able to prioritize applications as per the norms of Labor office and process the applications. Such system in turn leads to optimum use of available resources for achieving the targets and reduce the gap between educational institutions and labor market in Kingdom of Saudi Arabia (KSA).

A. Background

Kingdom of Saudi Arabia (KSA) being the 14^{th} largest country with a population of 28 million having a GDP of \$ 600.4 billion is having an overall literacy of 78% [1]. Since the day formal primary education started in 1930, the standards of education and their policies are being changed in KSA. The budget allocations towards education in KSA have been observed to be rising day by day with the raise in competition and challenges in this technological world. The education system in KSA is classified into Technical and Vocational Training Corporation (TVTC), Ministry of Civil Services, Ministry of Defense and Aviation, Ministry of Health and Saudi Commission for Health Specialties, Ministry of Higher Education and Royal Commission Jubail and Yanbu [1]. As a whole the government is giving highest priority for the education and largest share (≈ 25%) of budget is being allotted for education out of overall budget in KSA. The approximate amount of such budget of \$57.9 billion USD towards education is as large as any budget of worldwide countries. Such educational programs must be utilized and must be effectively useful for the students and citizens of KSA.

In the present market, most of the talent is being blindfolded due to lack of awareness of opportunities and appropriate information. The investment made by the government on education system is to be effectively utilized by transforming it into appropriate directions by proper communicating methodologies. The colleges and universities at various locations in KSA must be integrated to implement effective communication among the students and relevant information from the labor office.

B. Motivation

The role of this research will meet the requirements of government authorities to fulfill the needs of a job seeker in an efficient manner. Identifying the qualified student or right person for the job description and entitle the opportunities in a systematic manner so that the false happening can be reduced. At the same time the proposed method will be having the complete information of students from different universities and their academic credentials and their performance reports at various levels. Such information can be arranged properly and can be used as per the demands of the job providers/ business entities. In the present system communicating the information by means of emails, telephonic conversation, face-to-face interaction etc. is a hectic process and covering all students across the nation is a real challenging task. Such issues are being solved in the proposed method by integrating student information as a centralized communication process is established by this method. Managing the student information, prioritizing the student details according to labor office

regulations and employers will become an accessible task by using current method.

C. Related Work

Many universities in western countries have developed an integration system to communicate the student information with labor office and embassy for a systematic approach to deal with employment and regulations. The labor office will be able to prioritize the job seekers according to the requirements of the job seekers and promote the eligible candidates or students who graduated from different universities across the country. However, European Union (EU) countries are even sharing the student information among different nations to fulfill the demands of their markets and employers [2]. Most of the higher education systems (HEI) in EU (approximately above 90%) are tracking their students after graduation and their performance at work places in different job roles.

In Philippines, many universities are adopting the alumni tracking systems to provide a platform for the employers and the students to select suitable candidates and jobs respectively. For example, FAR Eastern University, Manila, is maintaining an "Alumni Relations and Placement Services (ARPS)" to provide a dynamic and responsive program for creating a strategic industry partnership to create efficient employment opportunities [3]. ARPS is providing the training sessions to the students for building the capabilities to work in real-time environments with different personalities and managements. These services are reducing the gap between the university training the industry demands by providing adequate training and information to the students who are seeking the jobs.

Martel tried to assess the practical techniques to design and study the impact of alumni over a period of ten years [4]. Tracking the alumni for ten years is a risky job; however, this intends to analyze significant changes in a systematic way. These changes could be positive or negative but such studies / techniques will help the educational institutions to understand the impact of a particular program in the student life to choose different careers or opportunities they have.

Similar studies were conducted by Tantawy *et al.,* [5] in Egypt. They conducted a detailed study on Information and Communication Technology (ICT) alumni from various Egyptian universities to track their employment and the places they are living in for various purposes. The complete process was conducted to reduce the gap between qualified students and job creation process in the country. They conducted this research based on the information available on social networking groups like Facebook, LinkedIn, Twitter, etc. A detailed distribution of alumni students working outside the Egypt are given below in figure 1 [5].

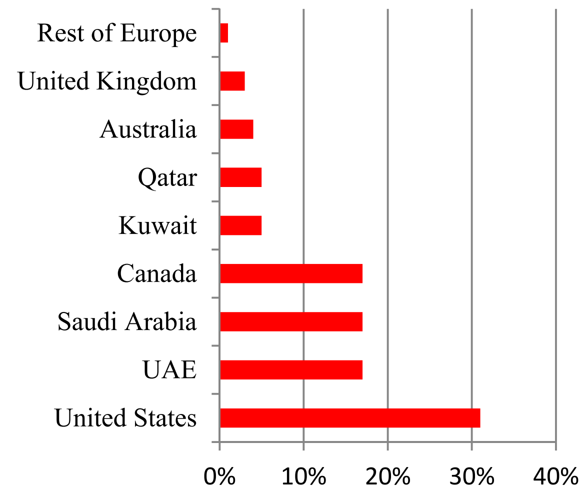


Figure 1. Alumni of Egypt working outside the country

Improper management of student alumni leads to a drastic situation in any nation towards increasing unemployment scenario for its own citizens.

II. Current System in Saudi Arabia

Tracking Alumni in Saudi universities are still at the initiation process only. Most of the alumni are working or progressing slowly due to poor maintenance and lack of understanding. Very few institutions are maintaining the alumni track record but the experiences or ideas of alumni members are considered very rarely in their academic development programs. Some of the alumni in SA are only using their websites as a source of latest job information with no regular updates, viz., http://alumni.ccdi-sorsogon.net/index.php. Such websites are maintained with poor and outdated information. They are independent and are not having adequate funding facilities to support. The information shared on such platforms is excluded with labor rules and the students seeking job will be unaware of exact information. Possibilities of students getting into legal problems after completing the graduation are more with no sufficient information about job roles or their impact on their lives. Due to these reasons in KSA the increased percentage of foreigners in labor forces has been observed as shown in figure 2 [6]. Apart from that, many Saudi students are not participating in most of the interviews due to low wages as compared to the government jobs as shown in figure 3 [6]. Having inadequate information with students is one major issue as discussed earlier is added with this new issue of wages to see large range of reduction towards participation for interviews due to low wages. The wages paid by private organizations are mostly suitable for foreigners but it does not suit local students.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

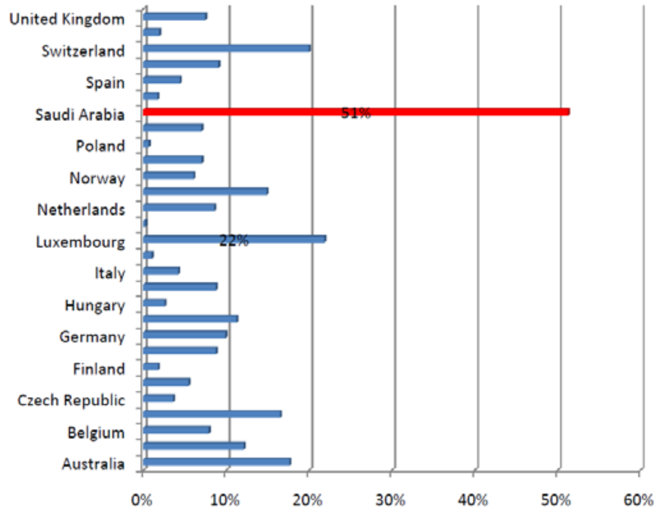


Figure 2. Increased Foreigners percentage in Labour Force

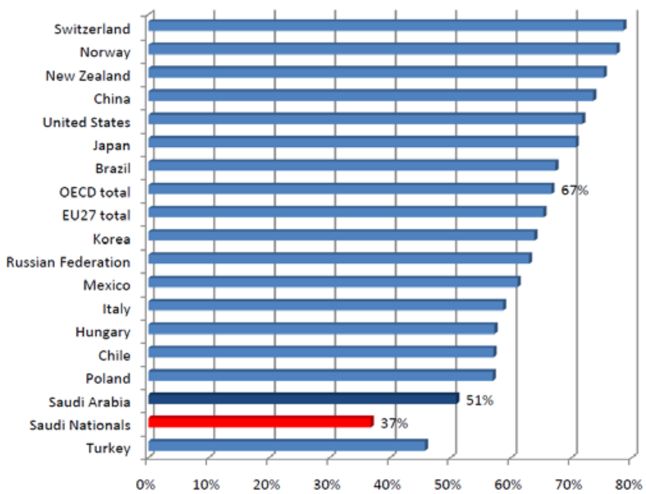


Figure 3. Reduction in Participation by Saudi Students for Interviews

The main purpose of alumni is to form an interaction between the graduated students and the students who are pursuing the education in the universities. Such interactions needs to explore the outside world, different work areas, market trends, demands, latest developments, critical technological aspects, etc. As a whole the job seeking candidates in universities need to have full information about present market and needs to get a direction to prepare for the forthcoming challenges in an effective manner. However, in most of the universities of KSA, the maintenance and tracking of alumni is very poor and rather it can be said that they are only for the sake of name.

The trend in KSA universities needs to change with present market requirements, increasing unemployment and to meet the social challenges in the society. A developed society needs to have a strategy of good education system with flexible information sources within the reach of its students. Most of the fresh graduates are unaware of the opportunities, labor rules, etc. With a dilemma the students will get distracted towards selecting a proper career after graduation. Recent reports from various industrial organizations revealed the importance of industry interaction with educational institutions for a better societal development.

Many methods and approaches are being followed by various nations to track the alumni. But in a country like Saudi Arabia the number of students visiting abroad is increasing day by day as per the reports of Institute of International Education [7]. The rate of students visiting to USA between 2006 to 2014 has increased at a rapid rate (+33%) as compared to any country after China (+36 %) as shown in figure 4 [8]. Increasing demands of population and qualified students across the country needs a systematic approach to tackle the employment issues carefully.

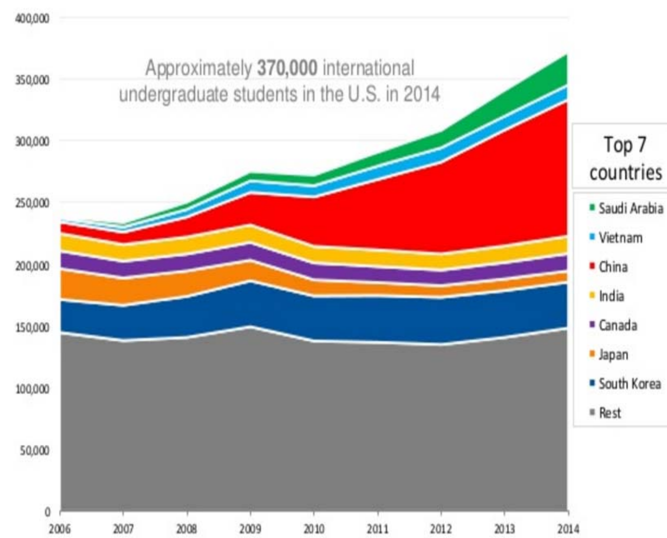


Figure 4. Increasing Trend of Students visiting USA for Higher Education

Many multi-national companies (MNCs) are keen to establish training and development units across the globe to develop a skilled workforce to sustain with the competition in the market. These MNCs are trying to collaborate with universities with a strategy to train the students and prepare local body entities to establish a good communication by creating strong alumni, which understands local language and customs. Managing the new students/ job seekers with the help of strong alumni will help these MNCs to improve the performance of their business goals and strategies.

III. PROPOSED ARCHITECTURE

The proposed architecture provides a solution to solve the unemployment problems in KSA in an efficient manner so that most of the students are benefited. This architecture consists of four main sections (figure 5) to be considered and is listed below:

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

A. Ministry of Education and Training Environment
B. Ministry of Labor Environment
C. Runtime Checker
D. Employers / Business Entities / Job Providers

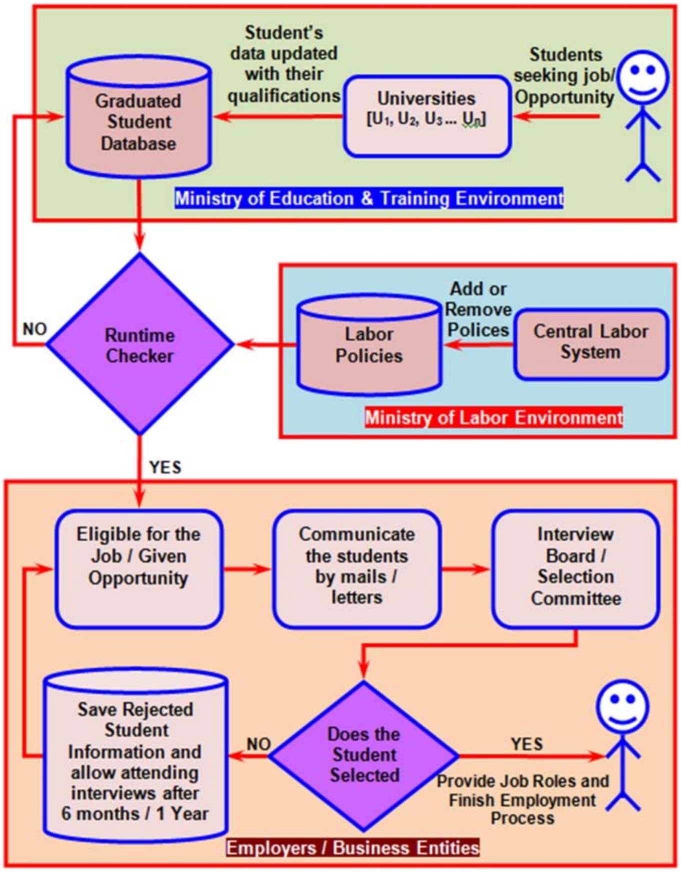


Figure 5. Proposed Architecture for Central Alumni System

A. Ministry of Education and Training Environment

Ministry of Education in KSA is having a lot of information about the students in different locations in different formats and procedures. A centralized information system is still missing from the Ministry of Education. An integrated database needs to be designed to collect the information of the students pursuing their education at different universities. Major steps proposed in this research toward establishing an integrated system includes the following areas to be covered:

1. A standardized format to cover all kinds of student activities in the universities
2. Categorize the Students based on their marks and performance in various aspects
3. Providing an easy access to the latest information related to the placement activities and job opportunities; and
4. Updating the graduate student database at regular intervals and also provide same information with labor office

Collecting the alumni information requires substantial resources for a good response from the students and job providers. Apart from gathering information the ministry also needs to identify the bench marking students and employers by adopting fair approaches in selection process. The ministry needs to ensure that the students are sent to genuine companies for their bright career and future.

B. Ministry of Labor Environment

Ministry of Labor needs to consider various student related scenarios and need to provide a clear list of regulation to employers and job seeking students. The information and amendments from the government must be updated with the universities and must ensure that these are available with various alumni database systems.

C. Runtime Checker

Runtime Checker needs to check the updates from graduated students database and labor policies database. This checker needs to collect the information and access the student data during runtime and based on labor policies, this data must be segregated. The eligible students' data must be forwarded to the Employers/ business entities for further process.

At the same time the rejected students list must be updated with the database along with various feedback reports explaining the reasons for their rejection.

D. Employers / Business Entities / Job Providers

The list of eligible students with different profiles will be allowed attend interviews at different organizations and business entities as per the student priority list. Communication is established between the companies and students for conducting and attending the interviews respectively. After a formal scrutiny process the interview boards or selection committees will deliver the selected and rejected students' list.

Selected students or job seekers will be provided with the details of company policies and regulations to be followed at work places. The rejected students' list will be saved into a separate database to ensure that these students are waitlisted for a period of 6 months/ one year based on the company's agreement with students.

IV. CASE STUDY: INTEGRATION OF UNIVERSITIES ALUMNI IN SAUDI ARABIA

In the whole process of developing the proposed architecture to maintaining and tracking alumni system strong database integration is essential. This integration process will help the large number of viewers (i.e. students, university administration, labor office, employers, etc). Various tools are available in market for integration and many companies are competing to provide such implementations. Mostly used integration methods involve [9]:

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

1. Extract Transform and Load (ETL)
2. Enterprise Application integration (EAI)
3. Enterprise Information Integration (EII)

The above mentioned integration tools are not discussed here as they are beyond the scope of present paper.

There are many universities in KSA and most of the information is being stored at different levels in different universities and places of entire kingdom. To establish the communication between two different universities a lot of time is being consumed in the process of communication and postal services. If it is related with confidential matters the information will be transferred as a matter of costly business. There are various aspects to be considered as discussed below:

A. *Initial Issues related with Data Collection and Integration*

All the universities must be having their own method of maintaining the student information and administration. Hence at the initial stage a standard method needs to be established and instructed to all universities across the KSA.

Collecting data, organizing the data, arranging the data and processing the data in the defined format will be a serious task.

B. *Design and Time Constrains*

Most of the students from KSA are studying across the nation and abroad needs to provide the information in a defined time. The information may not be similar according to the design as students studying abroad may have different formats of accessing the student performance. To standardize such issue a lot of time will be consumed.

Hence timely availability of the information, execution of a proper design and adopting the design to different formats of student assessments will be a challenging task.

C. *Scenario 1: Existing Recruitment Process*

In the present recruitment process the student needs to perform everything carefully right from his admission into a university to getting into a job. The student cannot be ignorant and needs to pay a careful attention towards the ultimate goal to reach to a job. In the whole process students needs to keep an eye on university rules and regulations to get qualified and also need to pay serious attention on the market trends. Apart from these the labor rules always tend to change based on the decisions of the government. The students need to have an eye over the changing labor rules, before getting into a serious issue. The existing recruitment process contains the following situations:

1. Students need to get into University by proper admission process and must be eligible for certain jobs requirement.
2. After completing the education students try to see the job advertisements.
3. Before attending the job interviews students need to fill application forms and need to know the rules and regulations for applying a job.
4. Also they have to understand the labor office rules and eligibility criteria (such as age, wages, working hours, etc).
5. Then student can apply for the job and the processing of an application will take place at an employer's place.
6. If the student is short listed means the student will get a call letter for applying the interview; otherwise a rejection letter with a feedback to disqualifying his/ her application.
7. The selected students will attend various stages of the interview and shortlisted candidates will be asked to stay for taking an appointment letter; and rejected candidates will be asked to leave the place.
8. The rejected candidates need to restart their search for job from step 2 until they get into a job.

In this process the students need to have a check on changing university rules and regulations and also need to update with current labor offices rules and market trends as an extra effort as shown in figure 6.

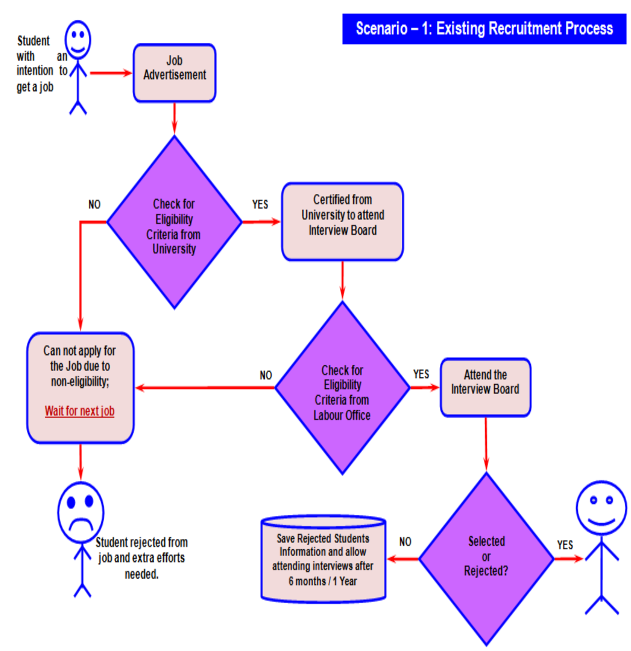


Figure 6. Existing Recruitment Process

D. *Scenario 2: Runtime Checker for Recruitment Process*

In the proposed recruitment process a unique method has been introduced by the author by providing a runtime checker. This runtime checker will try to check the updates from university rules and regulations and also from changing labor office rules and regulations. The students need not invest their time on updating themselves for the above mentioned as the runtime checker keeps an eye as a regular process of checking. The

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

proposed recruitment process contains the following situations:

1. Students need to get into University by proper admission process and must be eligible for recruitment processes.
2. The students need to ensure that their information is available with universities database and labor office database with all updates.
3. The employers willing to conduct interviews for candidates will check for the qualified students from the checker updates and calls the eligible candidates for interviews.
4. The selected students will attend various stages of the interview and shortlisted candidates will be asked to stay for taking an appointment letter; and rejected candidates will be asked to leave the place.
5. The rejected candidates need to wait for the next call letter from the employment office without wasting the time.

The formalized rules and requirements for the proposed architecture are shown in figure 7. The employment process by considering university rules and regulations along with the labor office rules are shown in the following cases. To establish these cases the above mentioned structure will be used.

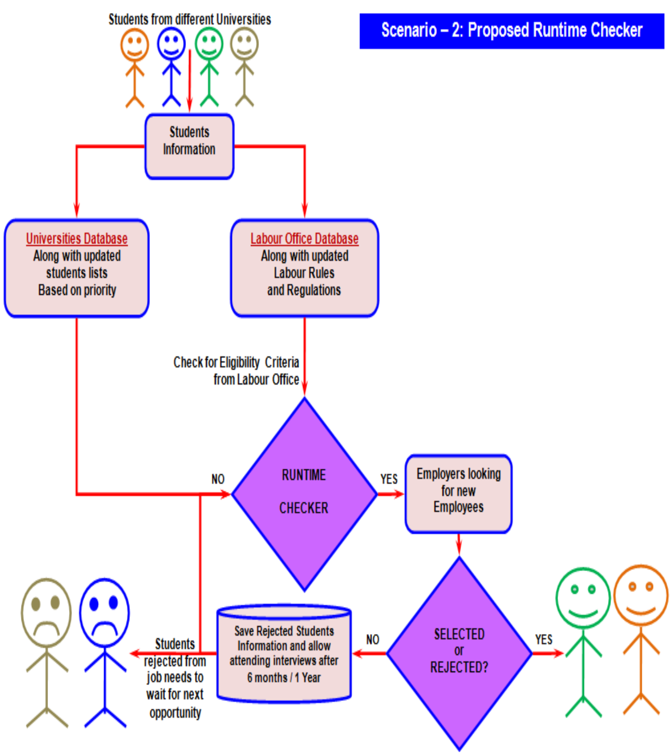


Figure 7. Proposed Recruitment Process

The student (S) minimum requirement (i.e. admission into university and passing the exams) to apply for a job in any company is formalized here as shown below:

$$Policy1 \triangleq \begin{pmatrix} fin(admissionUniversity) \wedge \\ (Student(S,\ Module) \wedge \\ PassExams(P, Module) \wedge \\ \wedge_{i=first\ year}^{t=course\ deadline}\ done\ (S,P,Submit) \end{pmatrix} \mapsto (Authorize^{+}(S,P,Apply\ for\ job))$$

After this the job requirements will be checked by the employer based on the courses students undergone in the university education; whether the student completed the course as required by the current position or anything else.

$$Policy2 \triangleq \begin{pmatrix} fin(EligibilityCriterion: Employer) \wedge \\ Course(C,\ Module) \wedge \\ Grades(G, Module) \wedge \\ \wedge_{i=0}^{i=Matching}\ done\ (C,G,Submit) \end{pmatrix} \mapsto (Authorize^{+}(C,G,Accept\ for\ job))$$

After this the labour rules must be checked for age, wages, nationality and background as follows:

$$Policy3 \triangleq \begin{pmatrix} fin(Rules, Regulations: LabourOffice) \wedge \\ (Age(A,\ Module) \wedge \\ Wages(W, Module) \wedge \\ Sex, M/F(X, Module) \wedge \\ Nationality(N, Module) \wedge \\ \wedge_{i=0}^{i=Matching}\ done\ (A,W,X,N,Submit) \end{pmatrix} \mapsto (Authorize^{+}(A,W,X,N,Accept\ for\ job))$$

The runtime checker will check the above three cases regularly and update the database and wait for the notifications from different employers for next recruitment session.

After above three cases, the students will be allowed to attend the interview; the selected students will finish the formalities of joining the companies. The rejected students will have to wait for the next interview process.

$$Policy4 \triangleq \begin{pmatrix} fin(Interview: Successful) \wedge \\ (Selected = Yes) \wedge \\ (FulfillFormalities = Yes) \wedge \\ (Consider = Yes) \end{pmatrix} \mapsto (Promote^{+}(C,G,Accept\ for\ job))$$

For the rejected students the information will be saved in the database and are supposed to try for the next job interview process:

$$Policy5 \triangleq \begin{pmatrix} (Interview = NotSuccessful) \wedge \\ (Selected = No) \wedge \\ (StudentInformation = SavedinDB) \wedge \\ (Consider = No) \end{pmatrix} \mapsto (Wait^{+}(\ Rejcted\ for\ job))$$

These policies will ensure that the students will not suffer much due to lack of information and are allowed to attend the

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

interviews with full information at a smooth processing system. For each job advertisement the monitoring system will ensure to check the university database and the labor rules for updates before allowing a student to apply for an interview. Such policies will help to reduce the overall time of students, universities and also the employers.

The algorithm (1) is based on the events required for the total recruitment process and various conditions involved in this process. Main task of this algorithm is to monitor the requirements of the employers from the opening of the activity to end of the activity and to apply the suitable policy to recruit the students as per the job requirement.

Algorithm 1 Check Job Requirement

Require: *Job[n], Job TM, Start T, Apply T, No of jobs, Job Period, Job State, T Left, Applied T, Selected Job, Major, Jadvertisement, JApply.*

Ensure: *n∈ No of Jobs*

```
Active is main state!, Inactive is initial state!
    FOR Jstate = Active DO
      (SelectingJ, ApplyingJ, AppliedJ, DeniedJ) // are substates!//
      DeniedJis initial substate!
        FOR Jstate = SelectingJ DO
        (getting available, Applied, Denied)  // are substates! //
        Getting available is initial substate!
        END FOR
    END FOR
REPEATE
    SelectedJ = Current Job( )
    IF n = SelectedJ THEN
        Jstate = Active
        Check Advertisement (Job [n]) //moving to selecting JState!  //
        Check availability (Job [n])   // moving to applying JState or
        Not Available JState  //
        Check Submission (Job [n]) moving to Applied!
        Jstate = In active   // Final State in this process!  //
    END IF
UNTIL Jstate = In active
```

The algorithm (2) will check the availability of the job during runtime. This process will continue until the total students are get selected from the alumni. This algorithm also checks whether the advertisement from an employer is suitable for the student's eligibility criterion or not.

Algorithm 2 Check Job Availability

Require: Job Policy for Availability

Ensure: Events Corresponds to the policy attributes

```
IF Jstate (Job[n]) ≠ unregistered ∧ got job ∧ unauthorized THEN
    Jstate (Job[n]) ← SelectingJ
    Javialable (Job [n])  ← getting availability //moving to suitable job
    state!  //
    Getting job policy (Job [n]) //Find applicable policy from the DB! //
RETURN Policy X
GET JEvents (Job[n])
RETURN Sequence of Event!
    T Left = CurrentMonth  // Year ( ) – Job Period //
    Check Policy F, D.Events
IF policy is satisfied THEN
    Javailability (Job[n]) ← Available
    JState (Job[n]) ApplyingJ   // Moving to ApplyingJ State! //
START M = CurrentM ( );
ELSE
    JAvailable (Job[n]) ← Denial
    Jstate (Job[n]) ← un authorized
END IF
END IF
```

V. Conclusions

The tracking of alumni becomes an essential aspect of Saudi Arabia due to increased competition and raising unemployment. Many students from other countries are visiting KSA for employment and are competing with the local students even for low wages. Due to which the local governing bodies are unable to maintain a peaceful employment policy as per the requirement of employers and labor rules of the government. Without maintaining an appropriate student alumni may lead to serious problem in the country if the jobs are not available for the own country students. Hence the proposed approach will surely help the government to adopt a systematic approach for maintaining and tracking the alumni of Saudi Universities.

References

[1] A. Morawski. The Educational System of Saudi Arabia. [Online] available at URL: <https://shelbycearley.files.wordpress.com/2010/06/education-in-saudi-arabia.pdf>, 2013.

[2] Gaebel, M., K. Hauschildt, K. Mühleck, and H. Smidt. "Tracking Learners' and Graduates' Progression Paths TRACKIT. Brüssel: EUA." (2012).

[3] President Report. Far Eastern University, 2015.

[4] M. Martel. Alumni Tracking: Challenges and Considerations of Designing an Innovative Study. December, 2012.

[5] Tantawy. R.Y., Farouk. Z., Mohamed. S., Yousef. A.H. Using Professional Social Networking as an Innovative Method for Data Extraction: The ICT Alumni Index Case Study. arXiv preprint arXiv:1410.1348. 2014.

[6] A. A. Omran. Unemployment in Saudi Arabia: A serious issue for the Saudi youth. London Business School. 2010.

[7] Institute of International Education. "International Students by academic Level and Place of Origin", 2014.

[8] D. Blough and M. Waxman-Lenz. "A New Era in Student Mobility: Emerging Opportunities for International Student Recruitment". Association of International Education Administrators, Annual Conference, Washington, DC, USA. February 15-18, 2015.

[9] P.K. Patro, P. Allen, M. Ramachandran, R. Morgan-Vane, and S. Bolton. "Databases and Information Systems Integration Using CALOPUS: A Case Study." In *ICEIS (1)*, pp. 200-207. 2006.

Authors Profile

Dr. Amr Jadi completed Doctorate degree from De Mountfort University and Masters from Bradford University, UK. The author is specialized in Software Engineering with an area interest in Early warning systems, Risk management and Critical Systems. The author is involved in various scientific activities of the Department of Computer Science and Software Engineering, University of Hail.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Secured Data Transmission in Wireless Sensor Networks

S. Suresh
Department of Information Technology
SRM University, Kattankaluthur Campus
Kancheepuram District, TamilNadu, India

Giridhar. R
Department of Information Technology
SRM University, Kattankaluthur Campus
Kancheepuram District, TamilNadu, India

Abstract—. **Security is one crucial requirement in Wireless Sensor network. To overcome this issue, security protocol called Didrip was developed for flat based network which allows for distributed data discovery and dissemination.**

But in terms of clustering approach which is most efficient one in terms of energy conservation, there are lot of security vulnerability i.e checking the cluster head for vulnerability to the network. In addition sensor nodes joining the cluster head during user joining phase is also not secure as the nodes can be vulnerable too. These two are most vulnerable security issues which are not addressed in existing security protocol of WSN including the one mentioned which is Didrip.

The above said problems for clustering approach in WSN are overcome with a Cluster-based Certificate Authority(CA) scheme which is combination of voting and Non-voting schemes towards detecting malicious node. We also use digital signature to sign all the nodes present in the network. These are simulated using standard network simulator ns-2 and results analysed in terms of packet delivery, network life time and energy efficiency.

Keywords- Didrip, WSN, CA, ns-2

I. Introduction

Wireless sensor networks (WSN) [1][2] are generally set up for gathering records from insecure environment. Nearly all security protocols for WSN believe that the opponent can achieve entire control over a sensor node by way of direct physical access. Wireless Sensor Networks are vulnerable to security attacks due to the broadcast nature of the transmission medium. Basically attacks are broadly classified into two categories i.e. active attacks and passive attacks [1][2]. Under passive attacks we have Monitor and Eavesdropping, Traffic Analysis, Camouflage Adversaries. In terms of Active attacks, we have Routing attacks, attacks on information in transit, selective forwarding, and Black hole/Sinkhole attack. Similar to passive and active attacks, there are also flooding attacks [1][2] in Wireless Sensor networks. Basically, via flooding attack, a malicious node/an attacker aim the exhaustion of the network resources (e.g. network bandwidth) as well as consuming the resources of an authentic network user (e.g. computational and battery power). Furthermore an attacker can influence the network performance, by hindering the proper execution of routing algorithm (in routing discovery phase). By Route Request (RREQ) flooding (or routing table overflow), it is possible for an attacker to send multiple RREQs to non-existing recipient in a very short period of time, using the Ad hoc On Demand Vector (AODV) protocol of Mobile Adhoc Networks (MANET). In other words the malicious node represents false (non-existing) routes to all authentic nodes within the network, preventing the creation of new actual ones and causing routing table overflow by the authentic users. The avalanche of RREQs all over the network leads to consummation of the battery power and the network bandwidth, causing Denial and Service attack.

As a countermeasure against the flooding attack, every network participant (actual authentic user or simply node) can compute and monitor the evaluation of all neighbors RREQ, and in case of outmatching of the RREQs' limit, which is preliminarily defined, the specific neighbor node comes with its ID in a blacklist. By this way, the authentic/actual node "knows" that it should not receive any RREQs from its neighbors recorded in its blacklist. Furthermore the efficiency of this countermeasure can be enhanced if the RREQ limit is not preliminarily defined (fixed), but is computed on hand

In Secure and Distributed Data Discovery and Dissemination in Wireless Sensor Networks [3]**,** a protocol named DiDrip was implemented. They are based on the centralized approach and only the base station can distribute data items. Such an approach is not suitable for emergent multi-owner-multi-user WSNs. Second, those protocols were not designed with security in mind and hence adversaries can easily launch attacks to harm the network. This research here allows the network owners to

 https://sites.google.com/site/ijcsis/
ISSN 1947-5500

authorize multiple network users with different privileges to simultaneously and directly disseminate data items to the sensor nodes. DiDrip consists of four phases, i.e., system initialization, user joining, packet pre-processing and packet verification. DiDrip consists of some flaws which is that the security protocol DiDrip was implemented for flat based system. Also in clustering approach cluster head is vulnerable to security. Secondly, User joining phase in clustering is not secure. Thirdly, Digital signature increase delay rate.

In terms of securing MANET, there are also lot of research been conducted in securing network. Certificate Revocation to Cope with False Accusations in MANET [4] play an important role in maintaining network security because attackers can freely move and repeatedly launch attacks against different nodes. By adopting certification systems, it becomes possible to exclude identified attackers from the network permanently by revoking the certifications of the attackers. So accordingly a certificate revocation scheme been proposed which revoke the certification of attackers in a short time with a small amount of operating traffic. By clustering the nodes and introducing multi- level node reliability, the proposed scheme can mitigate the improper certificate revocation due to false accusations by malicious users.

In another research [5], we build a clustering-based certificate revocation scheme, which outperforms other techniques in terms of being able to quickly revoke attackers' certificates and recover falsely accused certificates. To solve this problem, a new method been developed to enhance the effectiveness and efficiency of the scheme by employing a threshold based approach. In this method, node's accusation ability been restored and ensure sufficient nodes to accuse malicious nodes in MANETs. Extensive simulations show that the new method can effectively improve the performance of certificate revocation.

In another research in MANET, Cluster-Based Certificate Revocation with Vindication Capability for Mobile Ad Hoc networks[6] was developed. In this, we recover the warned nodes to take part in the certificate revocation process to enhance the accuracy. Also the threshold-based mechanism assesses and vindicates warned nodes as legitimate nodes or not before recovering them. The results demonstrate that the proposed certificate revocation scheme is effective and efficient to guarantee secure communications in mobile ad hoc networks

So in Wireless Sensor network, normally the clustering approach is used towards energy efficiency. So towards the clustering approach, the Cluster heads are selected according to the probability of optimal cluster heads determined by the networks. After the selection of cluster heads, the clusters are constructed and the cluster heads communicate data with base station. DiDrip is one of the security protocol implemented in flat based WSN which consists of four phases, i.e., system initialization, user joining, packet pre-processing and packet verification. Demerits of Didrip are that the Digital signature increase delay rate. Cluster head is vulnerable to network security and User joining phase is not more secure in Didrip. So accordingly we here developed a Cluster-based Certificate Authority scheme [7] for WSN which allows WSN nodes to be digitally signed by the Base station. This allows the WSN nodes to communicate with the cluster head and also if node moves from one cluster head to other, there is no issue towards exchange of keys towards secured transmission. In addition to cluster based certificate authority scheme towards communication, we also developed Voting and Non-Voting Scheme implementation for detecting the malicious nodes towards blacklisting and revocation of falsely accused nodes which is unique and novel in WSN. Voting and non voting scheme for malicious nodes been adopted in MANET where processing power and memory is not much of constraint comparing to WSN nodes.. Also our scheme can quickly revoke the malicious device's certificate, stop the device access to the network, and enhance network security. The rest of paper are organised as follows. Section II talks on literature review pertaining to security in WSN and MANET. Section III talks on system architecture and details pertaining to developed Cluster based Certificate authority scheme. Section IV talks on implementation and simulation analysis of Cluster based Certificate authority scheme using ns-2. Section V is the conclusion and Future work.

II. LITERATURE SURVEY

In this section we would be discussing some of the research work carried out in securing Wireless Network which are MANET and WSN.

A. *Certificate Revocation to Cope with False Accusations in MANET*

In Mobile Ad hoc NET works (MANETs), certification systems [4] are very important in maintaining network security. By adopting certification systems, identified attackers are excluded from network permanently by revoking certifications of attackers. Normally to identify attackers,

information about attackers are collected from nodes in network. Also in this mechanism, it becomes difficult to differentiate between valid and false accusations made by legitimate and malicious nodes respectively. As network size becomes large, the amount of traffic towards exchanging information about attackers and time to gather information about attackers increases. So towards this, Certificate revocation scheme been developed where attacker's certificate been revoked in short time with small amount of operating traffic. The improper certificate revocation on account of false accusation by malicious users can be mitigated by clustering nodes and introducing multi node reliability

B. A Study on Certificate Revocation in Mobile Ad hoc Network

In Certificate revocation, malicious node's certificate is revoked which denies nodes from all activities and isolate from network. In this research [5], clustering based certificate revocation scheme been developed which outperforms the other techniques towards revoking quickly attacker's certificate and also recover falsely accused certificates of nodes. However looking into the issues pertaining to certificate accusation and recovery, the number of nodes capable of accusing malicious nodes decreases over time. This ultimately leads to a situation where malicious nodes cannot be revoked at all. So towards giving solution to the problem mentioned, a threshold based scheme been developed towards restoring node's accusation ability and also ensuring sufficient normal nodes to accuse malicious nodes. The simulation results have shown that the new method of threshold based scheme improve the certificate revocation performance

C. Cluster-Based Certificate Revocation with Vindication Capability for Mobile Ad Hoc Networks

Mobile ad hoc networks (MANETs) are more vulnerable to security attacks on account of wireless and dynamic nature. Major challenge in MANET is to guarantee secured network services. So towards meeting the challenge in MANET, certificate revocation [6] is been employed to secure network communications. In this research, certificate revocation been focused towards isolating attackers from further participating in network activities. So toward quick and accurate certificate revocation, Cluster-based Certificate Revocation with Vindication Capability (CCRVC) scheme been employed. In this scheme towards improving reliability, warned nodes are recovered towards taking part in certificate revocation process and for enhancing accuracy, the threshold based mechanism employed towards assessing and vindicating warned nodes as legitimate nodes or not before recovering them. The numerical and simulation analysis have shown the performance of our scheme. Lastly the extensive results have shown that our proposed scheme is effective and efficient in guaranteeing secure communications in MANET.

D. DiDrip Security Protocol in Wireless Sensor Networks

In a research named Secure and Distributed Data Discovery and Dissemination in Wireless Sensor Networks [3], a protocol named DiDrip was implemented. They are based on the centralized approach and only the base station can distribute data items. Such an approach is not suitable for emergent multi-owner-multi-user WSNs. Second, those protocols were not designed with security in mind and hence adversaries can easily launch attacks to harm the network. This research here allows the network owners to authorize multiple network users with different privileges to simultaneously and directly disseminate data items to the sensor nodes. DiDrip consists of four phases, i.e., system initialization, user joining, packet pre-processing and packet verification. DiDrip consists of some flaws which is that the security protocol DiDrip was implemented for flat based system. Also in clustering approach cluster head is vulnerable to security. Secondly, User joining phase in clustering is not secure. Thirdly, Digital signature increase delay rate

III. CLUSTER BASED CERTIFICATE AUTHROITY IN WSN

Didrip [3] is one of the security protocol implemented in flat based WSN which consists of four phases, i.e., system initialization, user joining, packet pre-processing and packet verification. In MANET lot of security mechanism implemented using clustering approach. So we here have developed Cluster based certificate Authority Scheme [7] for Wireless Sensor networks where Digital Signature scheme used for signing all the nodes present in that network that helps in communication with cluster head even if the node moves from one network to other and thereby providing integrity . In addition to Cluster based certificate authority scheme for communication in wireless sensor network, we also have developed voting and nonvoting scheme similar to MANET which can quickly revoke the malicious device's certificate, stop the device access to the network, and enhance network security. The overall system architecture of Cluster based Certificate authority Scheme for wireless sensor node in Clustering approach is shown in Fig.1. The architecture got lot of components like route table maintenance, energy collector for selecting cluster head, Evidence collector for validating the certificate of wireless sensor node before communication with CH. All these are maintained in Route Manager along with the normal routing table information for routing data.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

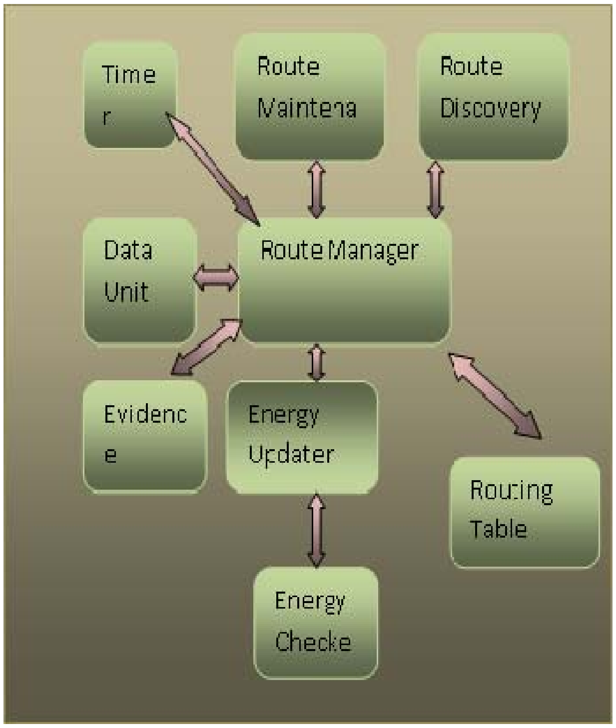


Figure 1 System Architecture of Certificate Scheme

A. Cluster Formation

Nodes cooperate to form clusters, and each cluster consists of a CH along with some Cluster Members (CMs) located within the transmission range of their CH. While a node takes part in the network, it is allowed to declare itself as a CH. In this model, if a node proclaims itself as a CH, it propagates a CH Hello Packet (CHP) to notify neighboring nodes periodically. The nodes that are in this CH's transmission range can accept the packet to participate in this cluster as cluster members. On the other hand, when a node is deemed to be a CM, it has to wait for CHP. Upon receiving CHP, the CM replies with a CM Hello Packet (CMP) to set up connection with the CH. Afterward, the CM will join this cluster; meanwhile, CH and CM keep in touch with each other by sending CHP and CMP. These are shown in Figure.2

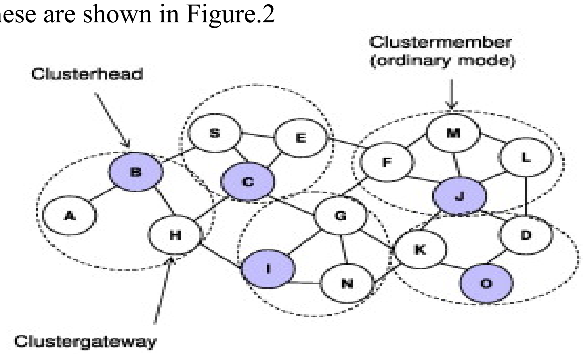


Figure 2 Cluster Formation

B. Certificate Authority

Before nodes can join the network, they have to acquire valid certificates from the Certificate Authority (CA), which is responsible for distributing and managing certificates of all nodes, so that nodes can communicate with each other unrestrainedly. The CA is also in charge of updating two lists, Warned List and Blacklist, which are used to hold the accusing and accused nodes information, respectively. Concretely, the Blacklist (BL) is responsible for holding the node accused as an attacker, while the Warned List (WL) is used to hold the corresponding accusing node. The CA updates each list according to received control packets. Note that each neighbor is allowed to accuse a given node only once. These are shown in Figure 3.

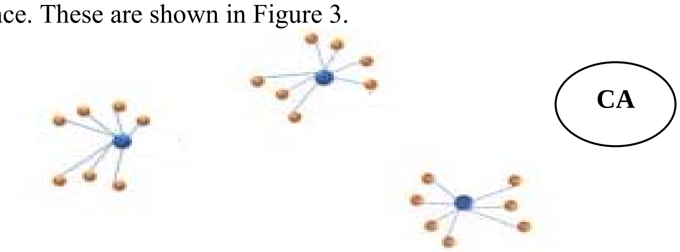


Figure 3 Certificate Authority

C. Node Classification

According to the behavior of nodes in the network, three types of nodes are classified according to their behaviors: legitimate, malicious, and attacker nodes. In our scheme, these nodes can be further classified into three categories based on their reliability: normal node, warned node, and revoked node. When a node joins the network and does not launch attacks, it is regarded as a normal node with high reliability that has the ability to accuse other nodes and declare itself as a CH or a CM. Moreover, we should note that normal nodes consist of legitimate nodes and potential malicious nodes. Nodes that are listed in the warning list are deemed as warned nodes with low reliability. Warned nodes are considered suspicious because the warning list contains a mixture of legitimate nodes and a few malicious nodes. Warned nodes are permitted to communicate with their neighbors with some restrictions, e.g., they are unable to accuse neighbors any more, in order to avoid further abuse of accusation by malicious nodes. The accused nodes that are held in the blacklist are regarded as revoked nodes with little reliability. Revoked nodes are considered as malicious attackers who are deprived of their certificates and evicted from the network. These are shown in Figure 4.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

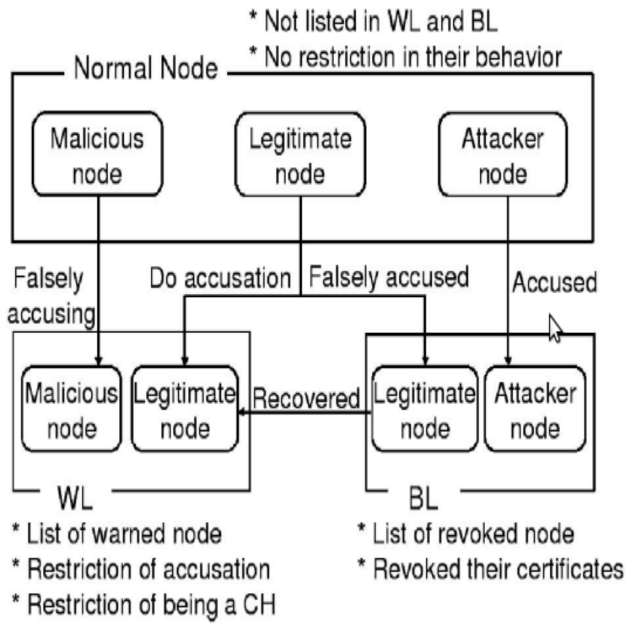


Figure 4 Node Classification

D. Certificate Revocation

To revoke a malicious attacker's certificate, we need to consider three stages

- Accusing
- Verifying
- Notifying

The revocation procedure begins at detecting the presence of attacks from the attacker node. Then, the neighboring node checks the local list BL to match whether this attacker has been found or not. If not, the neighboring node casts the Accusation Packet (AP) to the CA. Note that each legitimate neighbor promises to take part in the revocation process, providing revocation request against the detected node. After that, once receiving the first arrived accusation packet, the CA verifies the certificate validation of the accusing node: if valid, the accused node is deemed as a malicious attacker to be put into the BL. Meanwhile, the accusing node is held in the WL. Finally, by broadcasting the revocation message including the WL and BL through the whole network by the CA, nodes that are in the BL are successfully revoked from the network. These are shown in Figure 5.

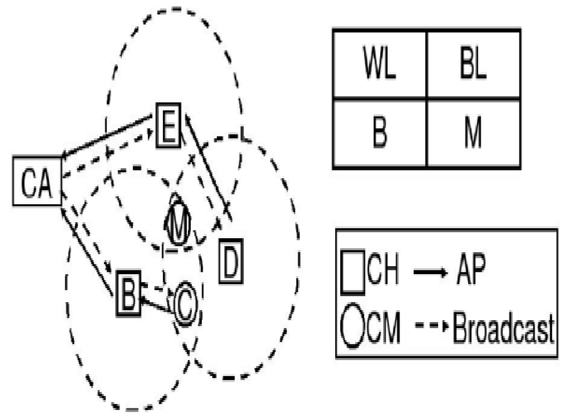


Figure 5 Certificate Revocation

E. Algorithm for Certificate Authroity Scheme

In our research, we have developed a security system with centralized unit. Centralized device can issue the certificate to all the user nodes. Certificate contains the hacker node and suspicious node information. These are shown in the form of Algorithm and same used in our implementation

- Initiate flooding attack in the network to eliminate attacker node.
- The attacker node will share hello message in a repeated manner which will flood the whole network and this will affect data transmission
- These repeated requests will get updated in the Routing Table of the other nodes present in that network as Ticks.
- In non-voting scheme adjacent nodes to the attacker node will complain to the CA about the node acting malicious based on the ticks.
- In voting scheme all the registered nodes present in the network will vote against the malicious node.
- Now the CA will compare the votes between voting and non-voting scheme
- Based on the votes CA will black list the particular node and a new route is found for further communication

IV. IMPLEMENTATION RESULTS AND ANALYSIS

The simulation of Didrip and Cluster based Certificate Authority scheme[7] for Wireless Sensor network security been carried out in ns-2. In addition, voting and nonvoting scheme in Cluster based Certificate Authority Scheme implemented towards detecting malicious nodes.

Figure 6 shows the clustering approach where cluster head selected for these 50 nodes by forming into groups which result in 7 cluster heads. Based on the cluster heads formed, a secret key is shared between cluster head and sensor nodes for communication as shown in Figure 7. In here we take a scenario where node 4 is cluster head and node 15 and 1 send

the request for certification for communication as shown in Figure 8. Once node 15 and 1 are certified as shown in Figure 9, data transmission occurs. Now the data transmission occurs between node 1 and node 15 through the gateway nodes 0 and 21 as shown in Figure 10. Figure 11 shows the data successfully transmitted between node 1 and node 15.

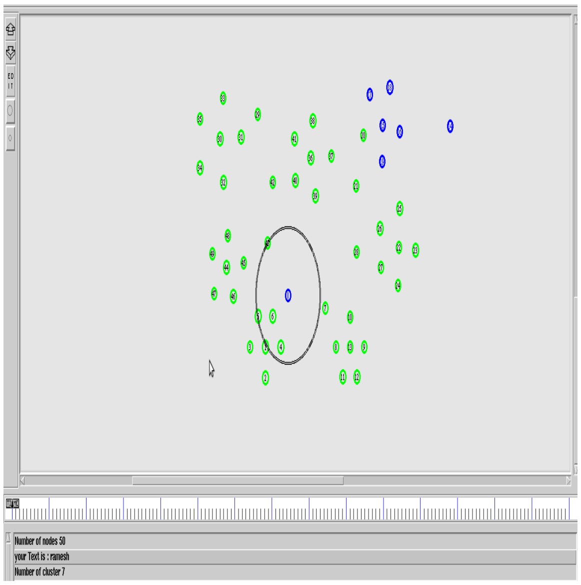


Figure 6 Cluster head Formation

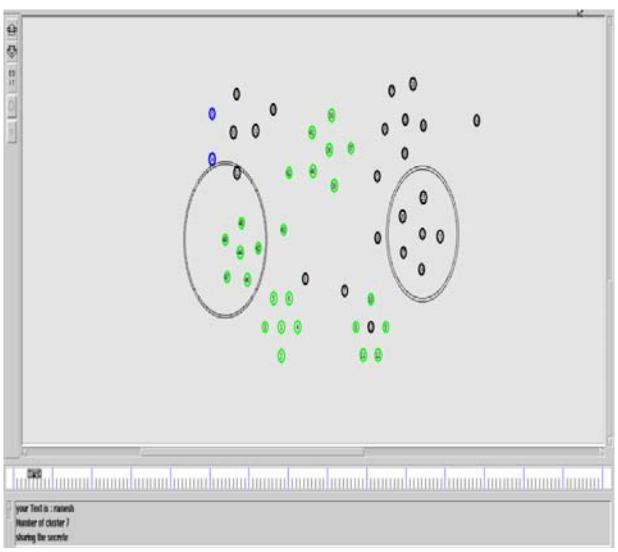


Figure 7 Secret Key Shared

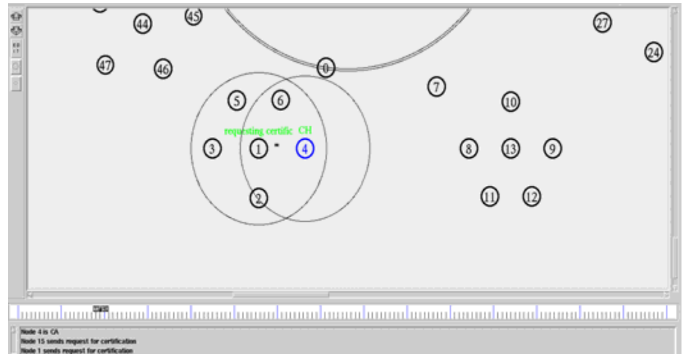


Figure 8 Certification Request

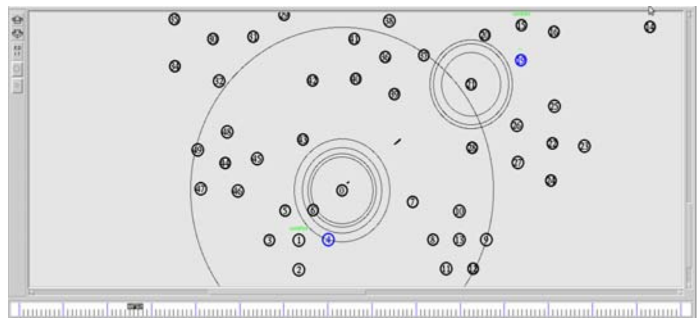


Figure 9 Data Transmission between Certified nodes

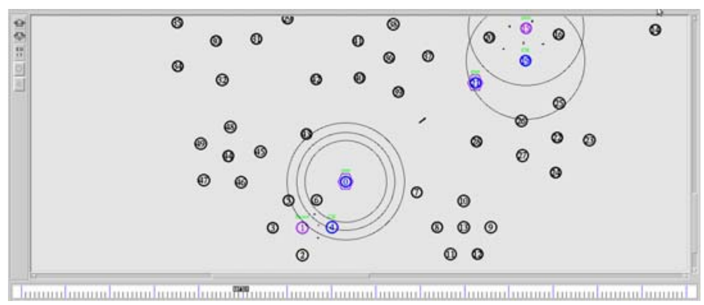


Figure 10 Data Transmission through gateway nodes

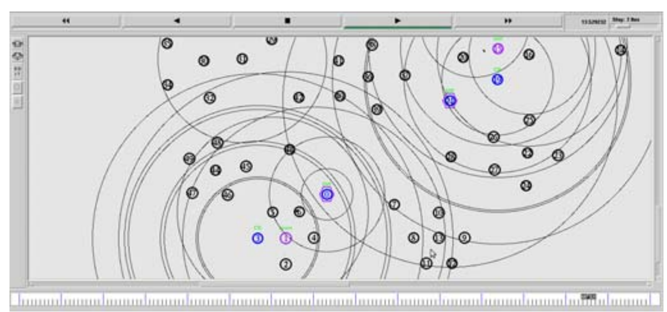


Figure 11 Data Transmission Success

Based on simulation of Didrip of sharing the key and certifying the nodes for data transmission through the gateway nodes, we see that the packet sent is 1150 and packet received is 1120 with good amount of packet loss coming to about 30. In addition the route overhead is also more which is 10. This is due to key sharing for every node wishing to transmit the data.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Also average end to end delay is 0.043 sec and packet delivery fraction is 99.02%. These are shown in Table 1.

Send packets	1150
Received packets	1149
Packet delivery function	99.9%
Average end to end delay (ms)	0.013s
No. of packets dropped	1
Route Overhead	0

Table 1 Didrip Clustering Analysis Results

In terms of Cluster based Certificate Authority scheme [7], cluster head selected from 50 nodes by forming into groups which result in 7 cluster heads. Now in Figure 12, we have created node 50 which is base station as Certificate authority and is shown in Red color. Its responsibility is to sign all the nodes towards communication. Then for all the group of sensor nodes, cluster head created as shown in blue color. In addition the nodes certified by CA are shown too.

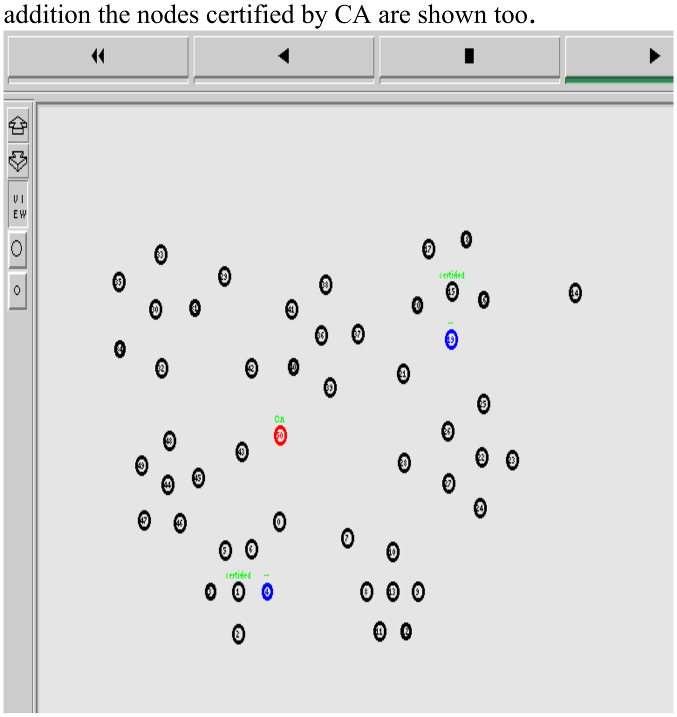


Figure 12 Certificate Authority Scheme

Send packets	1150
Received packets	1149
Packet delivery function	99.9%
Average end to end delay (ms)	0.013s
No. of packets dropped	1
Route Overhead	0

Table 2 Certificate Authority Analysis Results

In this method, the nodes wanting to communicate need not exchange the key and certificate request each time they want to communicate which creates unnecessary route overhead and also lot of packet loss as seen in in Didrip. So accordingly in simulating Cluster based Certificate authority scheme, each of the certified and digital signed nodes would communicate with the cluster head which verifies with the Base station which is node 50 towards communication. So even if the node moved from one Cluster head to other, there is no issue in communication as nodes are certified by Base station and no necessity to exchange key and certificate for every node communication which creates lot of overhead and packet delay. So accordingly we find that in Cluster head based certificate authority scheme, the packet sent is 1150 and packet received is 1149 resulting in minimal packet loss of 1 compared to Didrip. Also the route overhead is totally 0 and also end to end delay is 0.013s which is less compared to Didrip. Also the packet delivery fraction is 99.9 % which is nearly 100% compared to Didrip. These are tabulated in Table-2.

In terms of detecting malicious nodes using Certificate Authority Scheme, we have adopted voting and nonvoting scheme in Certificate Authority Scheme developed in Wireless Sensor Networks similar to MANET. The challenge in WSN is the processing power and memory which is big constraint as compared to MANET. In terms of Cluster based Certificate Authority scheme, cluster head selected from 50 nodes by forming into groups which result in 7 cluster heads. Now in Figure 13, we have created node 51 which is base station as Certificate authority and is shown in Red color. Its responsibility is to sign all the nodes towards communication. Then for all the group of sensor nodes, cluster head created as shown in blue color which is shown in Figure 14. In addition

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

the nodes certified by CA are shown too. Then the route request is shared between the nodes for initiating communication between the source and the destination node as shown in Figure 15. Then the reply message for the route request is shared in the network and it will get updated in the routing table is shown in Figure 16. Finally the data transmission takes place between the source and the destination node as shown in Figure 17. In Figure 18, the attacker node floods the network with hello message and by using nonvoting scheme the attacker node makes a false accusation against the legitimate node and CA puts that particular node in the warning list. Then the voting scheme comes into play for all the registered nodes present in the network who will vote against the malicious node based on the ticks created due to repeated Hello message sharing as shown in Figure 19. Now the comparison is done by CA based on the votes from both voting and nonvoting scheme and the CA shares the attacker nodes certificate and that particular node is added into the black list as shown in Figure 20. Finally the route is changed due to malicious node detection and the data transmission is performed as shown in Figure 21.

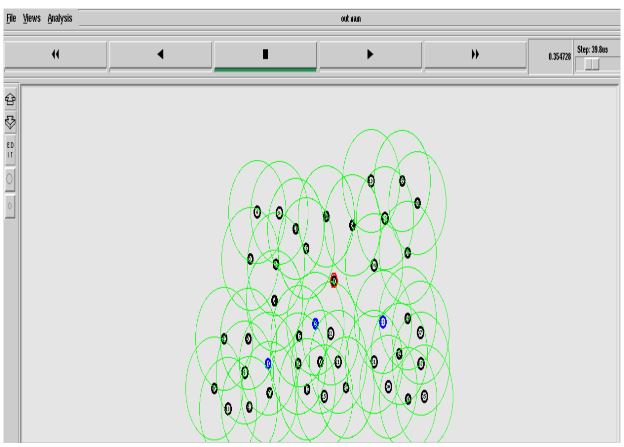


Figure 13 Hello message sharing

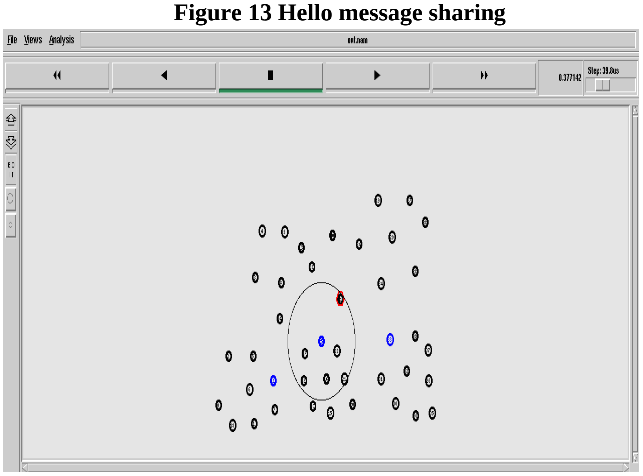


Figure 14 CH advertisement

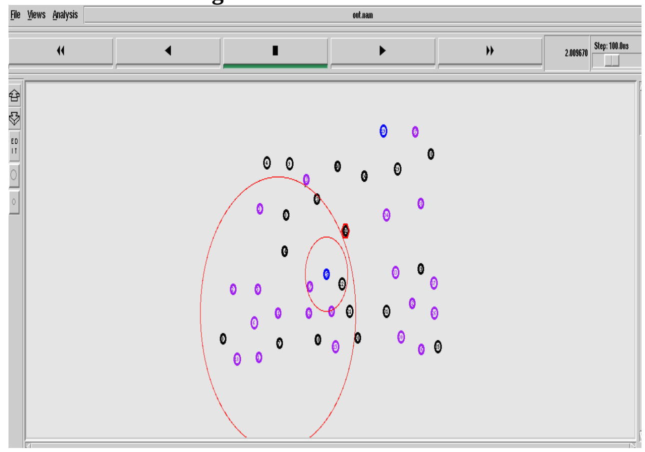


Figure 15 Route request sharing

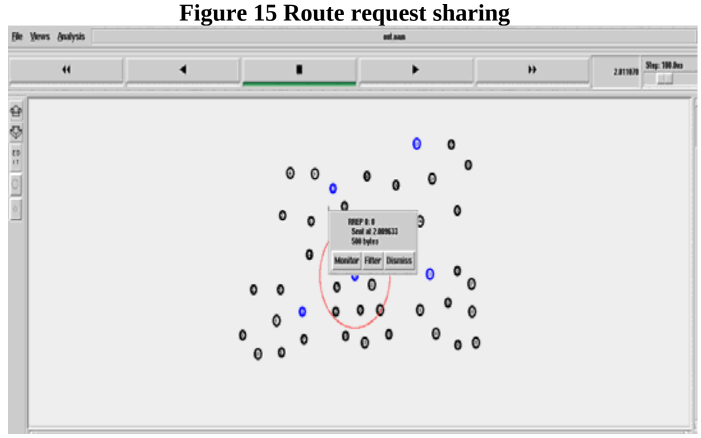


Figure 16 Reply message sharing

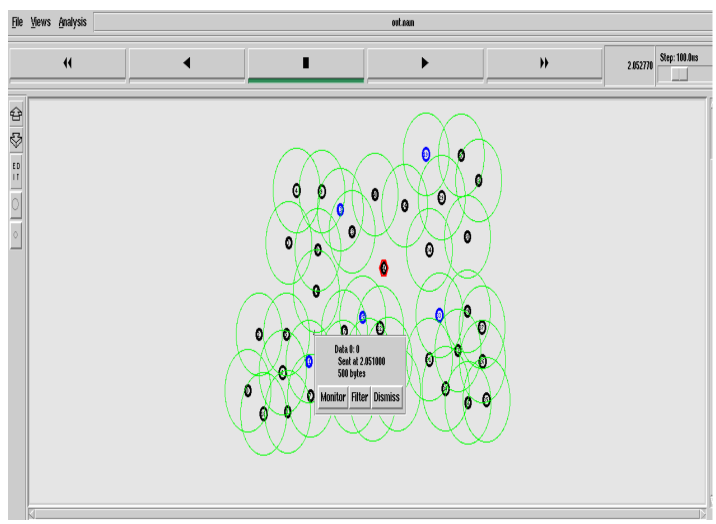


Figure 17 Data Transmission

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

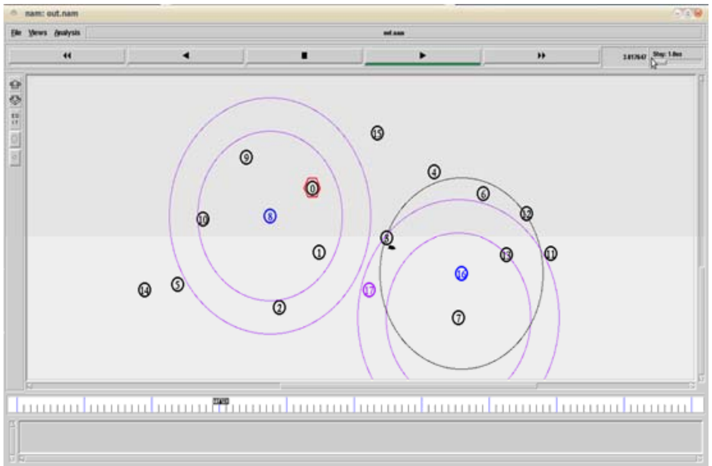


Figure 18 Flooding Network - Attacker

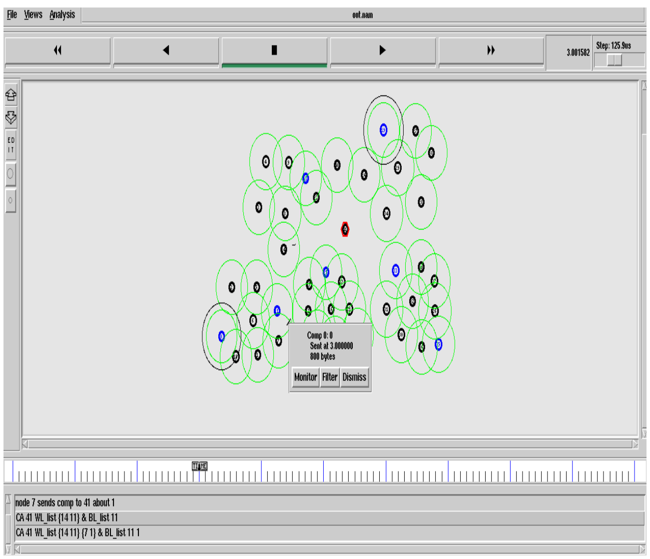


Figure 19 Complains about malicious node

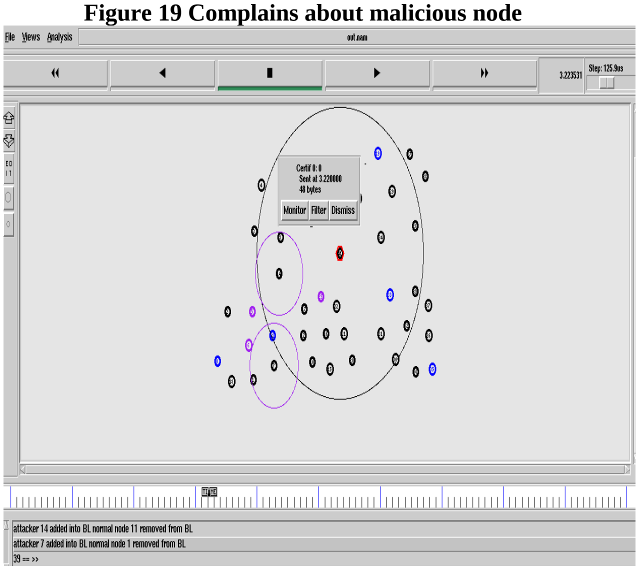


Figure 20 CA Certificate Sharing

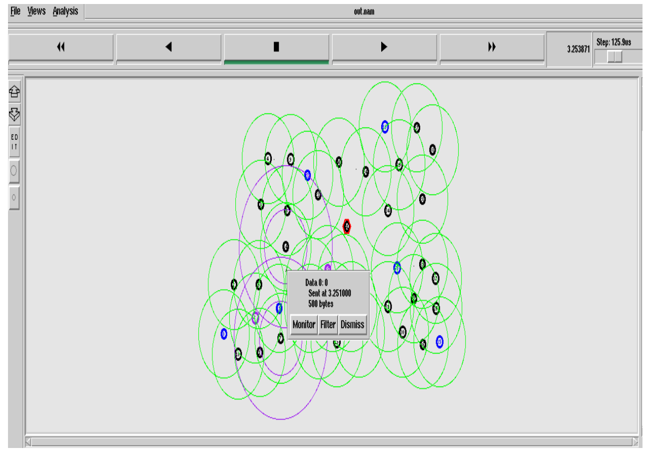


Figure 21 Route changed - Malicious node detection

Now in terms of simulation analysis towards detecting the malicious nodes, the packet delivery factor in Certification Authority scheme is high when compared with Didrip protocol as shown in Figure 22. Green color represents implementation of CA and the red color represents Didrip in Figure 22..

During the time of flooding attack, the packet drop is more in CA which after eliminating the malicious node, the packet delivery factor increases drastically as compared with the old security protocol i.e Didrip . Due to intense key exchange process in Didrip, the life time of the network is less as shown in red color. But in CA, it is less complex and highly efficient nature in terms of eliminating malicious node which is shown in green color. Also consumption of energy is high in Didrip protocol compared to Certificate Authority for providing security. That is energy efficiency in CA is much better compared to Didrip which is shown in Figure 23 as Green and Red color.

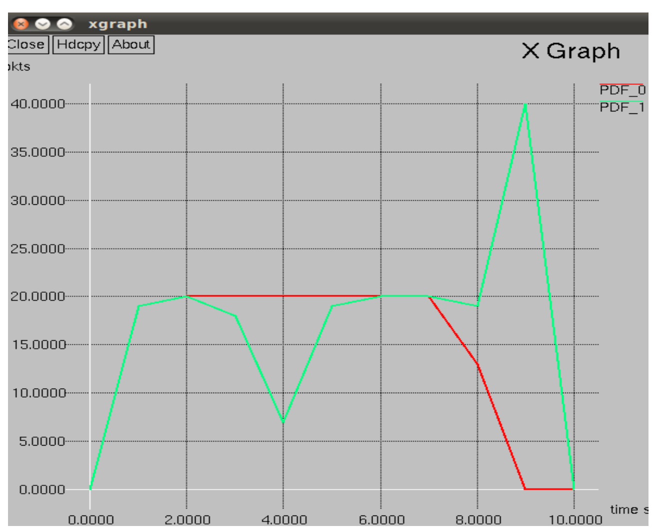


Figure 21 Packet delivery - Didrip and Certification authority

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

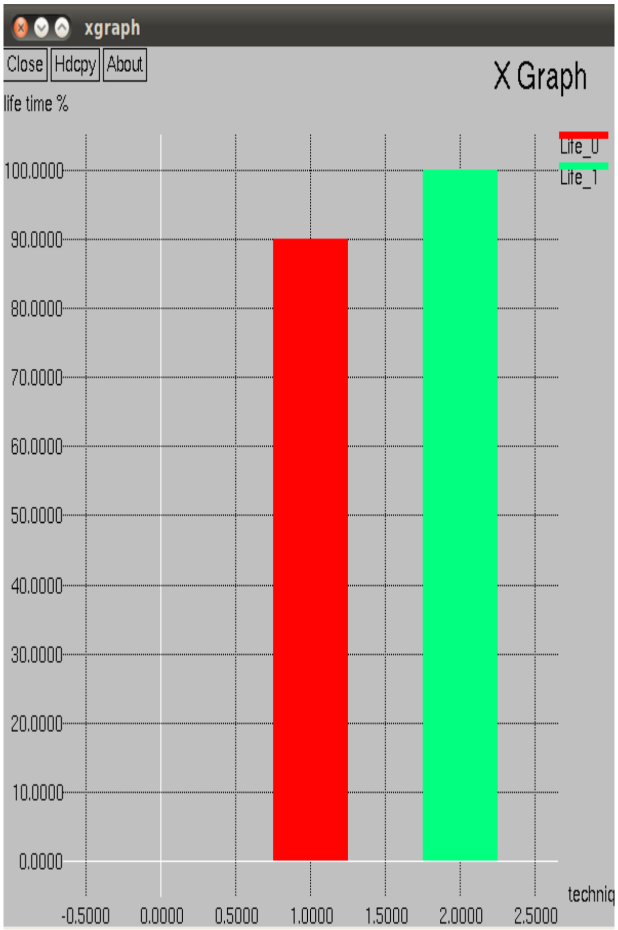


Figure 22 Network lifetime- Didrip and Certification Authority

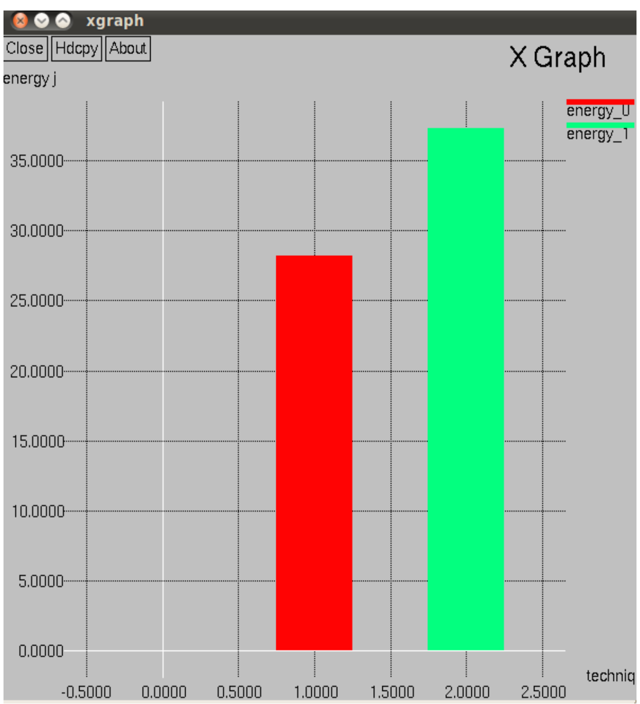


Figure 23 Energy efficiency – Didrip and Certificate Authority

V. CONCLUSION & FUTURE WORK

Security in Wireless Sensor network is a very crucial requirement considering the different types of application it is being deployed today. Quite amount of research carried out in terms of clustering approach towards energy conservation in wireless sensor networks towards data dissemination and communication.

One of the security protocol implemented in flat based WSN is Didrip towards secured data dissemination. In terms of MANET, lot of research been carried employing clustering approach towards detecting malicious nodes, revoking the certificate and so forth.

So accordingly, we in this research have developed a cluster-based certificate Authority scheme combined with the merits of both voting-based and non-voting based mechanisms to revoke malicious certificate and solve the problem of false accusation. The scheme can revoke an accused node based on a single node’s accusation, and reduce the revocation time as compared to the voting-based mechanism. In addition, we have adopted the cluster-based model to restore falsely accused nodes by the CH, thus improving the accuracy as compared to the non-voting based mechanism. Particularly, we have proposed a new innovative method to release and restore the legitimate nodes, and to improve the number of available normal nodes in the network. In doing so, we have sufficient nodes to ensure the efficiency of quick revocation. The extensive results have demonstrated that, in comparison with the existing method Didrip, our proposed scheme is more effective and efficient in revoking certificates of malicious attacker nodes, reducing revocation time, and improving the accuracy and reliability of certificate revocation. We have proven that based on cluster head selection scheme, we have improved the energy saving. These have been simulated and analyzed using ns-2 simulator. In future, we would be working more towards network load and Qos issues towards deployment of Certificate authority in Wireless Sensor network. Also there is need to look into black hole and worm hole attack in Wireless sensor networks by employing voting and nonvoting scheme. Last but not the least the research can be extended with Internet of things pertaining to security too.

REFERENCES

[1]. Mohammad., and M. Imad, , Handbook of Sensor Networks: Compact Wireless and Wired Sensing System, CRC Press, 2005

[2]. Toh, C-K . Adhoc Mobile Wireless Networks: Protocols and Systems. Prentice Hall, 2001

[3]. C. Sammy, Daojing, G. Mohsen., Y. Haomiao, Z. Boyang, “Secure and Distributed Data Discovery and Dissemination in Wireless Sensor Networks”, IEEE Transactions on Parallel and Distributed Systems, 26(4):1129-39, 2015.

[4]. L. Wei, N. A. Hirok, “Study on Certificate Revocation in MANETS”, IEEE International Conference on Communications (ICC), pp.1-5, 05-09 June 2011.

[5]. P. Kyul, N. Hiroki., A. Nirwan, K. Nei, “Certificate Revocation to cope with False Accusation in Mobile Adhoc

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Networks”, 71st IEEE Vehicular Technology Conference, pp.1-5, 16-19 May 2010

[6]. L.Wei, N. Hiroki, A. Nirwan, Y. Jie, “Cluster-based Certificate Revocation with Vindication Capability for MANETS”, IEEE Transactions on Parallel and Distributed Systems, 24(2) : pp234-42, 2013.

[7]. Giridhar, R and Suresh, S. “Cluster based Certificate Authority Scheme in Wireless Sensor Networks”, Proceedings of 2016 IEEE International Conference on Wireless Communication, Signal Processing and Networking, Chennai, India, March 23-25, 2016, pp. 382-286

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 6, June 2016

A multi-step method to calculate the equilibrium point of the Continuous Hopfield Networks: Application to the max-stable problem

Mohammed EL ALAOUI [1], Karim EL MOUTAOUAKIL [2] and Mohamed ETTAOUIL [1]
[1]Modelling and Scientific Computing Laboratory
University Sidi Mohammed ben Abdellah, Fez, MOROCCO
[2]National school of applied sciences Al-Hoceima (ENSAH) BP 03, Ajdir Al-Hoceima

***Abstract*— The Continuous Hopfield Networks (CHN) is a neural network tools which can be used to solve many problems like auto-memory and optimization problems. The dynamics of the CHN is described by differential equations system which is hard to solve analytically. That is why, the researchers use the Euler Cauchy method to calculate the CHN equilibrium point. Unfortunately, this method suffers from several problems, especially quality of the decision for a large step, sensibility to the slope function parameters and to the initial conditions. In this work, we use the well-known multi-step numerical method called Adams–Bashforth method, which is strong in terms of stability and performance, to calculate the equilibrium point of the CHN associated with the max stable problem. This method introduces an intermediary step to improve the Euler Cauchy method precision. The experimental results show that the (CHN+Adams-Bashforth) method produce a large max stable sets in comparison with the (CHN+Euler-Cauchy) method.**

***Key-Words:* - Continuous Hopfield Networks, Euler Cauchy method, Adams–Bashforth method, max-stable problem.**

I. INTRODUCTION

Hopfield has shown its ability to handle a wide variety of optimization problems, such as the traveling salesman problem, and also has been used as an associative memory in the image processing field. In 1982, The Hopfield model of Neural Networks took the name of its inventor, John Hopfield that will strengthen later the neural network by a new model [1], [4].

In contrast to the other neural networks, in this Hopfield network, the information does not flow in a single direction but flows only from input to output. This network has shown that the use of a continuous transfer function ensures existence stable state if weight matrix is symmetric with zero diagonal. The energy function of this network is monotonically decreasing with time, which makes convergence to a local minimum can be guaranteed.

The dynamics of the Hopfield networks are formulated as analytically differential equation. This gives the numerical discretization neural network scheme in which each neuron state changes with iteration. Much of the work on this type of network has received remarkable importance in its application in different domains [5],[7],[24]. On the other hand, the discretization of a continuous system does not always benefited from mathematical tools to be applied. Many researchers have worked on the discretization Hopfield network by the method of Euler. However, this method suffers from several problems, especially quality of the decision for a step size. In order to overcome this problem, in this work, we discretize the Hopfield network with Adams-Bashforth method [21]; this method will show its power to calculate the equilibrium point for each iteration. In this work, we use the well-known multi-step numerical method called Adams–Bashforth method which is strong in terms of stability and performance. In fact, this method introduces an intermediary step to improve the Euler Cauchy method precision [19]. The discretization of the Hopfield neural network with the Adams–Bashforth method makes it more robust than the classical network. The proposed system is used in this work to solve the max-stable problem using the CHN to evaluate the performance of this approach by comparing with the classical Hopfield.

This paper is organized as follows: In section II, we present an introduction of the continuous Hopfield network. The equilibrium point of CHN and the local error of the second order Adams-Bashforth method are discussed in the section III. In section IV, the maximum stable set problem is modelized as a 0-1 quadratic program. Some numerical results using both (CHN+Adams-Bashforth system) and (CHN+Euler-Cauchy system) are presented in this section V. Finally, section VI provides a conclusion and future work.

II. THE CONTINUOUS HOPFIELD NETWORK (CHN)

The model proposed by John Hopfield in 1982 is always presented as a powerful tool for solving many problems as the traveling salesman problem, linear programming problems, graph coloring problems, processing image, constraint satisfaction problems [7]. Due to the diversity of use of this model in various ways, this model has attracted many

researchers attention to work with this model. The Hopfield networks are fully connected networks, the weight matrix is symmetric $T_{ij} = T_{ji}$ the strength of the connection from neuron j to neuron i. Each neuron i has an offset bias of i_i^b .The dynamics of the CHN is represented by the differential equation:

$$\frac{du}{dt} = -\frac{u}{\tau} + Tx + i^b$$

Where u, x and i_b will be the vectors of neuron states, outputs and biases. The output function $x_i = g(u_i)$ is a hyperbolic tangent, which is bounded below by 0 and above by 1. $g(u_i) = \frac{1}{2}(1 + \tanh(\frac{u_i}{u_0})) \quad where \quad u_0 > 0$

Where u_0 is a parameter used to control the gain (or slope) of the activation function. If, for an input vector u_0 , a point u^e exists such that $u(t) = u^e \quad \forall t > t_e$, for some $t_e \geq 0$, this point is called an equilibrium point of the system defined by the differential equation [16]. That point equilibrium point is also called the stable point system. The Hopfield model can be written as a Lyapunov function, so this model is stable and decreasing system over time. The evolution of each step gives a trajectory that converges to an equilibrium point, because the Lyapunov function provides the possibility of finding a local minimum. Hopfield showed that, if matrix T is symmetric, then the following Lyapunov function exists [2],[6],[28]:

$$E(x) = -\frac{1}{2}x^t Tx - (i^b)^t x + \frac{1}{\tau}\sum_{i=1}^{n}\int g^{-1}(v)dv$$

The CHN can solve any combinatorial problems, which seeks to minimize an objective function: $E(x) = -\frac{1}{2}x^t Tx - (i^b)^t x$

The main idea of this Lyapunov technique is in each step is stable and converges one of the local minima for any combinatorial problem [3]. In this way, the output of the Hopfield network is seen as a solution for many combinatorial problems.

Consider the following quadratic assignment problem, with n variables and m linear constraints:

$$(P)\begin{cases} Min \quad \frac{1}{2}x^t Qx + q^t x \\ subject\ to \\ \qquad Ax = b \\ \qquad x_i \in \{0,1\} \quad i = 1,....,n \end{cases}$$

To solve the quadratic programming (P) using the Continuous Hopfield Networks, the following sets are needed:

H is a set of the Hamming hypercube: $H \equiv \{x \in [0,1]^n\}$

H_C is a set of the Hamming hypercube corners : $H_C \equiv \{x \in H : x_i \in \{0,1\}, \quad i = 1,....,n\}$

H_F is a set of feasible solutions: $H_F \equiv \{x \in H_C : Ax = b\}$

Thus the process for a given instance (n,m,Q, q, A,b), some conditions must be established on the problem so that its equilibrium points can be associated with local minima of the optimization problem, with m is the number of constraints.

An energy function must also be defined by: $E(x) = E^0(x) + E^R(x) \quad \forall x \in H$ Where:

$E^0(x)$ is directly proportional to the objective function of the problem.

$E^R(x)$ is a quadratic function that not only penalizes the violated constraints of the problem, but also guarantees the feasibility of the solution obtained by the CHN . This function must be constant $\forall x \in H_F$ and an appropriate selection of this function is crucial for correct mapping.

This energy function was introduced to overcome the problem observed with the energy functions used by other authors, including Aiyer [8] and Brandt et al. [23].

In this paper, our goal is to solve the maximum stable set problem by a proposed new approach.

In this case, the next step is to discretize the Hopfield network with a new method called Adams-Bashforth method. And in the second step we present a modelization of the maximum stable set problem as a quadratic 0-1 programming. From this model, implementation step becomes easy and general.

III. The Equilibrium Point of CHN and Local Error of the Adams-Bashforth Method

Recently, continuous Hopfield networks (CHN) are used to solve very interested combinatorial problems like travelling salesman problem[26], graph coloring problems, placement of the electronic circuits problems, maximum stable set problem, constraint satisfaction problem and Optimization of the Kohonen Network Architectures Using the Continuous Hopfield Networks[14],[18],[23].

The dynamic of the CHN is characterized by the flowing differential equation that takes the general form:

$$\frac{du}{dt} = f(u)$$

Where $f(u) = -T \times \tanh(u) - I$

The discretization Hopfield network is often given by the numerical method of Euler method, which is defined by the following equation.

$$u_{n+1} = u_n + h \times f(u_n)$$

This method is highly sensitive to initial condition and step-size. In addition, the Euler method produces local solutions which are not enough good. To overcome this problem, we use in this work the second order Adams-Bashforth multi-step method. The idea of this method is to take different time steps for two components for to achieve the target accuracy, the components are integrated using larger step sizes. The large step sizes are at their turn integrated multiple of the small step

sizes.

Euler method is part of an Adams–Bashforth family [21]. To derive Adams–Bashforth formulas, notice that:

$$u(t_{n+1}) = u(t_n) + \int_{t_n}^{t_{n+1}} f(t, u(t))dt$$

The approximation to the integral is obtained from polynomial interpolation at the points:

$$(t_n, f_n), \ldots, (t_{n-k+1}, f_{n-k+1})$$

For some integer k, and for k = 1, Adams-Bashforth order 1 is Euler's method: approximates the integral by the area of a rectangle whose base has length $(t_{n+1} - t_n)$ and whose height is $f(t_n, u_n)$:

$$u_{n+1} = u_n + \int_{t_n}^{t_{n+1}} f(t, u(t))dt \approx u_n + f(t_n, u_n)(t_{n+1} - t_n)$$

The figure 1, explains geometrically this approximation.

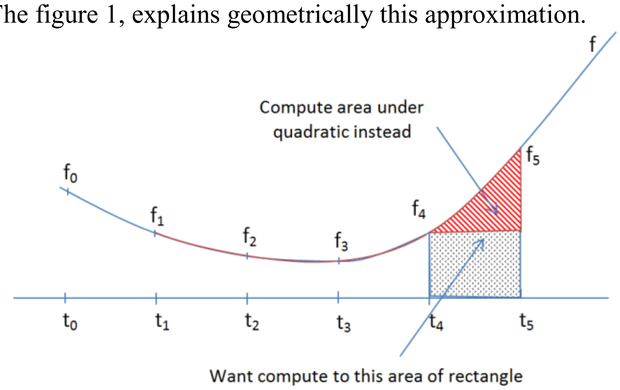


Figure 1: Adams-Bashforth method with k=1.

For Adams–Bashforth with k = 2, the interpolation of the integral is done by a polynomial of degree 1, with $p(t_n) = f_n, \quad p(t_{n-1}) = f_{n-1,}$

The figure 2, explains geometrically this approximation.

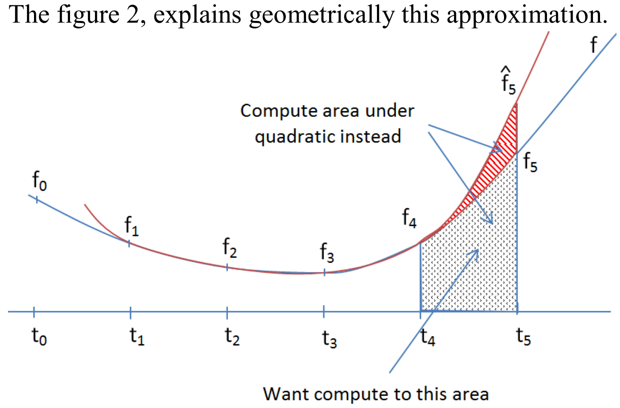


Figure 2: Adams-Bashforth with k=2

So letting $h_n = t_{n+1} - t_n$ and $h_{n-1} = t_n - t_{n-1}$, we obtain:

$$u_{n+1} = u_n + \int_{t_n}^{t_{n+1}} f_{n-1} + \frac{f_n - f_{n-1}}{t_n - t_{n-1}}(t - t_{n-1})dt$$

$$= u_n + h_n f_{n-1} + \frac{f_n - f_{n-1}}{h_{n-1}} \frac{(h_n - h_{n-1})^2 - h_{n-1}^2}{2}$$

By setting $h_n = h_{n-1}$, we obtain the Adam-Bashforth scheme:

$$u_{n+1} = u_n + \frac{h}{2}(3f_n - f_{n-1})$$

To determine the error of each order of the Adams–Bashforth family, one can call for the Taylor series if u is smooth enough [20]; see the table 1.

Table 1: Order and local error by Adams-Bashforth

Some Adams-Bashforth	Order	Local Error
$u_{n+1} = u_n + hf_n$	$k=1$	$\frac{h^2}{2}u^{(2)}(\eta)$
$u_{n+1} = u_n + \frac{h}{2}(3f_n - f_{n-1})$	$k=2$	$\frac{5h^3}{12}u^{(3)}(\eta)$

Here we have two parameters, step size h and order k, are used in control the size of the local error.

Practical methods for solving the differential equations use such estimates for the local error to determine whether the current choice of step size h is adequate.

The following example demonstrates the effectively of the method described above. In order to show the effectiveness of Adams-Bashforth method, we compare the local error for each iteration. The example has been used for comparison:

u'=u+1, x0 =0, u(x0)=1;
u(x)=2e^x-1 is the exact solution

Figure 3, contains the results obtained by two methods for stepsize h=0.1. Again, a comparison with the results of Adams-Bashforth method and Euler method shows that Adams-Bashforth gives comparable accuracy. The following figure shows the errors obtained for each of two methods:

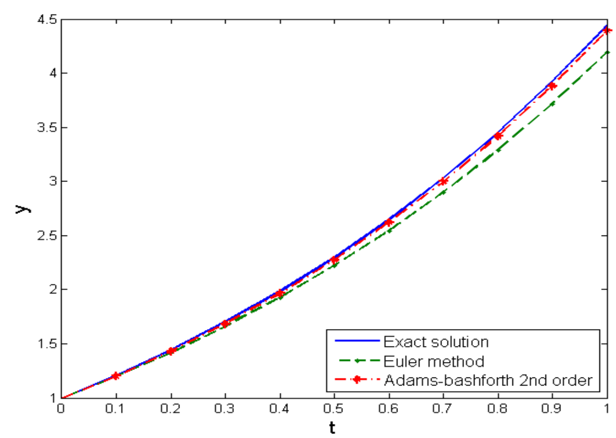


Figure 3: approximation by Adams-Bashforth and Euler method

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

The results are shown in figure 3. There, the relative error ε rel: and the number of iteration. Upon comparing the error of the second-order Adams-Bashforth and Euler method, we notice that the local truncation error is smaller for the Adams-Bashforth method; this reflects the effectiveness of the Adams-Bashforth method. In front of the results were observed the importance of Adams-Bashforth for calculating approximate solution of optimization problem. In this work, we use the Adams-Bashforth of order 2 to solve the Max-Stable problem and we compare the obtained results to those obtained with the Adams-Bashforth of order 1(Euler method) [22].

IV. Continuous Hopfield network for the Max-Stable Problem (MSP)

In this part, we construct an adequate Continuous Hopfield Network for the max-stable problem. To this end, we express the (MSP) in terms of 0-1 a quadratic program. Basing on this latter, we introduce our own energy function; then we use the hyper plan method to select adequate parameters to obtain a feasible stable set [12][13]. First, we define the MSP and we discuss the main proposed work to solve this problem.

A. The max-stable problem

Given an undirected graph G=(V, E) with $V = \{v_1, v_2, ..., v_n\}$. A stable set of a graph G is a set of nodes S with the property that the nodes of S are pairwise non adjacent. The Maximum Stable Set Problem (MSSP) consists of finding a stable set in graph G of maximum cardinality $\alpha(G)$. A side from its theoretical interest, the MSSP problem arises in applications in information retrieval, experimental design, signal transmission, and computer vision [25]. The stable set problem is NP-hard in the strong sense, and hard even to approximate. The MSSP problem can be solved using polynomial time algorithms for special classes of graphs such as perfect graphs and t-perfect graphs, circle graphs and their complements, claw-free graphs, and graphs with long odd cycles [27], [11] and [15]. But, the existence of a polynomial time algorithm for arbitrary graphs seems unlikely.

Different approaches have been discussed in the literature to solve the maximum stable set problem exactly. An implicit enumeration technique of Carrahan's and Pardalos's [13], computational results for different stable set linear programming relaxations have been reported by Gruber and Rendl [14], an effective evolution of the tabu search approach is presented in the original work of Friden, Hertz and de Werra [15]. The MSSP problem can be solved via the Continuous Hopfield Network (CHN).

B. 0-1 a quadratic program for the max-stable problem

To solve the MSSP problem via the CHN, it must be expressed as an assignment problem with a quadratic constraint.

Let $S \subset V$ be a stable set of nodes. For each node v_i of the graph G, we introduce the binary variables x_i such that:

$$x_i = \begin{cases} 1 & if \quad v_i \in S \\ 0 & \quad Otherwise \end{cases}$$

Two adjacent nodes v_i and v_j cannot be in the set S: $(v_i, v_j) \in E \Rightarrow x_i x_j = 0$ The constraints can be aggregated in a single one:

$$h(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij} x_i x_j = 0$$

$$With \quad b_{ij} = \begin{cases} 1 & if \quad (v_i, v_j) \in E \\ 0 & Otherwise \end{cases}$$

The objective function of the mathematical programming model is:

$$f(x) = -\sum_{i=1}^{n} x_i \quad .$$

Consequently, the MSSP problem can be expressed in the following algebraic form:

$$(QP)\begin{cases} Min \quad f(x) = -\sum_{i=1}^{n} x_i \\ subject\ to \\ \qquad h(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij} x_i x_j = 0 \\ \qquad x \in \{0,1\}^n \end{cases}$$

C. The CHN for the Max-stable problem

The main objective of this section is to construct an adequate CHN to solve the maximum stable set problem (MSSP). Firstly, we begin by formulation of energy function associated with this MSSP problem. Then, we select a convenient parameters setting of this function [9], [10]. The formulation of this energy function for maximum stable problem is done as follows:

$$E^0(x) = -\alpha\sum_{i=1}^{n} x_i + \frac{1}{2}\phi\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij} x_i x_j + \gamma\sum_{i=1}^{n} x_i(1 - x_i)$$

We determine the weights and thresholds as follows:

$$\begin{cases} T_{i,j} = -\phi b_{ij} + 2\delta_{ij}\gamma \\ i_i^b = \alpha - \gamma \end{cases} \qquad (1)$$

with $\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$ is the Kronecker symbol.

The parameters ϕ, γ and α must be chosen so that the Hopfield network equilibrium point associated with the MSSP is realized. The parameter-setting procedure is obtained from the partial derivative of the energy function:

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$\frac{\partial E^0(x)}{\partial x_i} = E_i(x) = -\alpha + \phi\sum_{j=1}^{n} b_{ij}x_j + \gamma(1-2x_i)$$

The parameters-setting are determined by the hyper plane method [9]. Before treatment, some conditions are necessary to simplify the determination these parameters: $\phi > 0, \gamma > 0$.

To minimize the objective function, we impose the following constraint: $\alpha > 0$.

The coming constraint is necessary for system stability and which is obtained by the following equation $x \in H - H_C:\quad T_{i,i} = 2\gamma \geq 0$. Such as it was one constraint for the maximum stable set problem, we obtain: $H_C - H_F = \{x \in H_C \,/\, h(x) > 0\}$.

Let $x \in H_C - H_F$, in this case, two adjacent nodes v_i and v_j are in the stable set S , then $x_i = x_j = 1$ and therefore the value x_i will decrease if $E_i^0(x) \geq \varepsilon \quad where \quad \varepsilon > 0$. The following constraint is obtained: $-\alpha + \phi - \gamma \geq \varepsilon$.

All of these constraints will display the following result:

$$\begin{cases} \alpha > 0, \quad \phi > 0, \quad \gamma \geq 0 \\ -\alpha + \phi - \gamma = \varepsilon \end{cases}$$

A feasible solution could be the following:

$$\begin{cases} \alpha > 0, \quad \phi > 0, \quad \gamma \geq 0 \\ -\alpha + \phi - \gamma = \varepsilon \end{cases} \qquad (2)$$

Finally, the weights and thresholds Hopfield can be found based on the parameters of pre-treatment.

D. The proposed algorithm

Basing on the prosed Continuous Hopfield Network, for the max stable problem, the Adams-Bashforth method and using the equations 1 and 2, we propose the following algorithm:

Algorithm : (Adams-Bashforth discretization of the Hopfield network)

Input data

- -The graph G= (V, E);
- -The weight matrix and bias vector are Calculated from the equation (1);
- -The parameters α, ε *and* ϕ are positive real and γ is calculated from the system (2);
- -The Adams-Bashforth step h fixed to a small value.
- -The stopping criterion is MaxIter or a small non negative real eps.

Out put

Vector of binary elements.

Step 0.

Initialize randomly the neurons states

Step 1.

While the stopping condition is false, do Steps 2-6.

Step 2. Perform Steps 3-5.

Step 3. Choose a unit at random.

Step 4. Change activity on selected unit:

$$u_{n+1} = u_n + h/2 \times (3f_n - f_{n-1})$$

Step 5. Apply output function

$$v_n = 1/2 \times [1 + \tanh(u_n / u_0)]$$

Step 6. Check stopping condition.

As this algorithm converges rapidly to a local minima, we can turn it several time staring form several initial state; at the end we chose the best solution.

V. Simulation Results

In the present work, we showed the practicality of our approach in a series of experimentations to solve the max stable set problem. The evaluation instances are given to DIMACS Challenge [17]. These graphs were presented as test problems for aims solving the maximum clique problem. For these graphs, we tested each instance at the end of applying our approach to the maximum stable set problem. This implementation was done by using language Java and personal computer environment with an Intel CPU of Core i5 and 4 GB of RAM.

Calculate randomly generated initial states: $x_i = 0.55 + 10^{-5}t$

Where t is a random uniform variable in the interval [-0.5,0.5].

We choose the parameters:

$\alpha = 1.0250$, $\varepsilon = 10^{-6}$ and $\gamma = 0.7$; the parameter ϕ was computed from the equation $\phi = \alpha + \gamma + \varepsilon$

The results are supplied in table 2.

Table 2 : Computational results of the instances

graph	*V*	*|E|*	*α(G)*	*CHN Euler*	*CHN Adams*
				$\alpha_1(G)$	*$\alpha_2(G)$*
brock200_2	200	9876	12	11	11
brock200_4	200	13089	17	9	12
brock400_4	400	59765	33	6	9
brock800_2	800	208166	24	12	18
gen200_p0.9_44	200	17910	44	27	36
gen400_p0.9_55	400	71820	55	38	53
hamming8-4	256	20864	16	16	16
hamming10-4	1024	434176	40	40	40
keller4	171	9435	11	7	9
Keller5	776	225990	27	10	23
p_hat300_1	300	10933	8	8	8
p_hat300-3	300	33390	36	31	31
p_hat700-1	700	60999	11	11	11
p_hat700-2	700	121728	44	--	--
C125.9	125	6963	34	26	26
C250.9	250	27984	44	37	44
C1000.9	1000	450079	68	34	34
MANN_a27	378	70551	126	72	123

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

-- : the repetition of the unstable point.

α1(G): the size of the stable set obtained by CHN combined with the Euler Cauchy method.

α2(G): the size of the stable set obtained by CHN combined with the Adams-Bashforth method.

In order to obtain these results the machine needed 200 steps with Adams-Bashforth method and 200 steps with Euler method.

This table shows that the result is better when using the Hopfield network incorporated by the Adams-Bashforth method. In fact, the (CHN+Adams-Bashforth) system produce a large max stable sets in comparison with the (CHN+Euler-Cauchy) system.

In addition, upon comparing the error of the second-order Adams-Bashforth and Euler method, we notice that the local truncation error is smaller for the Adams-Bashforth method; this reflects the effectiveness of the Adams-Bashforth method. In front of the results were observed the importance of Adams-Bashforth for calculating approximate solution of optimization problem.

VI. Conclusion

In this work, we have used the well-known multi-step numerical method called Adams–Bashforth method to calculate the equilibrium point of the CHN associated with the max stable problem.

In order to confirm the practical effectiveness of this method, many simulations have been carried out, the graphs were taken from the 2nd DIMACS Challenge. The simulations results showed that the (CHN+Adams-Bashforth) method produce a large max stable sets in comparison with the (CHN+Euler-Cauchy) method. Such results are obtained thanks to the intermediary step, which permit to improve the Euler Cauchy method precision. The future work for this research is to use the (CHN+Adams-Bashforth) method to solve some well-known combinatorial problems such as the traveling salesman problem, linear programming problems, graph coloring problems, processing image and constraint satisfaction problems.

References

[1] H.Asgari, Y .S. Kavian, A. Mahani, A systolic architecture for Hopfield neural networks, Procedia Technology, Vol.17, 2014, pp. 736–741.

[2] G. Joya, M.A. Atencia, F. Sandoval, Hopfield neural networks for optimization: study of the different dynamics, Neurocomputing, 2002, pp. 219-237.

[3] Amanda C. Mathias, Paulo C. Rech, Hopfield neural network: The hyperbolic tangent and the piecewise-linear activation functions, Neural Networks, 2012, pp. 42-45.

[4] Tariq Samad, Paul Harper, High-order Hopfield and Tank optimization networks, Parallel Computing, 1990, pp. 287-292.

[5] Rachid Sammouda, Nuru Adgaba, Ameur Touir, Ahmed Al-Ghamdi, Agriculture satellite image segmentation using a modified artificial Hopfield neural network, Computers in Human Behavior, 2014, pp. 436-441.

[6] Rong Long Wang, Zheng Tang, Qi Ping Cao, A learning method in Hopfield neural network for combinatorial optimization problem, Neurocomputing, 2002, pp. 1021-1024.

[7] J.J. Hopfield, D.W. Tank, Neural computation of decisions in optimization problems, Biological Cybernetics, 1985, pp. 1-25.

[8] A.H. Gee and S.V.B. Aiyer, and R.W. Prager, An analytical framework for optimizing neural networks, Neural Networks, 1993, pp. 79-97.

[9] P.M. Talaván, J. Yàñez, The generalized quadratic knapsack problem, Neural Networks, 2006, pp. 416-428.

[10] J. Håstad, Clique is hard to approximate within $n^{1-\varepsilon}$, Acta Math,1999, pp. 105-142.

[11] C. Mannino and A. Sassano, An exact algorithm for the maximum cardinality stable set problem. Computational Optimization and Applications, 1994, pp. 243-258.

[12] K. Tatsumi, Y. Yagi, and T. Tanino, Improved projection Hopfield network for the quadratic assignment problem, Proceedings of the 41st SICE Annual Conference, 2002, pp.2295-2300.

[13] R. Carrahan and P. M. Pardalos, An exact algorithm for the maximum clique problem, Operations Research Letters, 1990, pp. 375-382.

[14] G. Gruber and F. Rendl, Computational experience with stable set relaxations, SIAM J Opt, 2003, pp. 1014-1028.

[15] C. Friden, A. Hertz, and D. de Werra, Stabulus: a technique for finding stable sets in large graphs with tabusearch, Computing, 1989, pp.35-45.

[16] P.M. Talaván and J. Yàñez,A continuous Hopfield network equilibrium points algorithm, computers and operations research, 2005, pp.2179-2196.

[17] DIMACS. The Second DIMACS Implementation Challenge, ftp://dimacs.rutgers.edu/pub/challenge/graph/benchmarks/ clique/.

[18] M. Ettaouil, M. Lazaar, K. El moutaouakil, K. Haddouch, A New Algorithm for Optimization of the Kohonen Network Architectures Using the Continuous Hopfield Networks, WSEAS TRANSACTIONS on COMPUTERS, 2013.

[19] J.C. Butcher, Numerical methods for ordinary differential equations in the 20th century, Journal of Computational and Applied Mathematics, 2000, pp. 1-29.

[20] Shichang Ma, Yufeng Xu, and Wei Yue, Numerical Solutions of a Variable-Order Fractional Financial System, Journal of Applied Mathematics, 2012.

[21] Shaban Gholamtabar, Nouredin Parandin, Numerical solutions of second-order differential equations by Adam Bashforth method, American Journal of Engineering Research, 2014, pp. 318-322.

[22] J.C. Chiou, S.D. Wu, On the generation of higher order numerical integration methods using lower order Adams-Bashforth and Adams-Moulton methods, Journal of Computational and Applied Mathematics, 1999, pp. 19-29.

[23] R.D. Brandt, Y. Wang, A.J. Laub, and S.K. Mitra, Alternative networks for solving the travelling salesman problem and the list-matching problem, proceedings of the International Conference on Neural Networks, 1988, pp. 333-340.

[24] Ferhat Kurtulmuş, İsmail Kavdir, Detecting corn tassels using computer vision and support vector machine, Expert Systems with Applications, 2014, pp. 7390–7397.

[25] Marc Demange, Tınaz Ekim, Bernard Ries, Cerasela Tanasescu, on some applications of the selective graph coloring problem, European Journal of Operational Research, 2015, pp. 307-314.

[26] A. Fischer, F. Fischer, G. Jäger, J. Keilwagen, P. Molitor, I. Grosse, Exact algorithms and heuristics for the Quadratic Traveling Salesman Problem with an application in bioinformatics, Discrete Applied Mathematics, 2014, pp. 97-114.

[27] Yerim Chung, Marc Demange, The 0–1 inverse maximum stable set problem, Discrete Applied Mathematics, 2008, pp. 2501-2516.

[28] F.H. Clarkea, Yu.S. Ledyaevb, R.J. Stern, Asymptotic Stability and Smooth Lyapunov Functions, Journal of Differential Equations, 1998, pp. 69–114.

An Event Grouping Based Algorithm for University Course Timetabling Problem

Velin Kralev, Radoslava Kraleva, Borislav Yurukov
Department of Informatics, South West University "Neofit Rilski", Blagoevgrad, Bulgaria

***Abstract* — This paper presents the study of an event grouping based algorithm for a university course timetabling problem. Several publications which discuss the problem and some approaches for its solution are analyzed. The grouping of events in groups with an equal number of events in each group is not applicable to all input data sets. For this reason, a universal approach to all possible groupings of events in commensurate in size groups is proposed here. Also, an implementation of an algorithm based on this approach is presented. The methodology, conditions and the objectives of the experiment are described. The experimental results are analyzed and the ensuing conclusions are stated. The future guidelines for further research are formulated.**

Keywords – university course timetabling problem; heuristic; event grouping algorithm

I. Introduction

The University Course Timetabling Problem (UCTP) is an optimization problem and has been widely explored for the last 55 years. For the first time the key aspects of this problem were presented in [1]. In order to solve a UCTP a finite number of events $E = \{e_1, e_2, ..., e_n\}$ synchronized in time and fixed on a timetable that consists of a finite number of time slots $T = \{t_1, t_2, ..., t_k\}$ is needed. The arrangement of the events must be done in such a way that it satisfies the finite number of hard constraints (C_h) and violates the fewest possible ones from a finite number of soft constraints (C_s). A timetable is acceptable when it meets all hard constraints and is better than another one when it violates fewer soft constraints [2].

The UCTP is NP-hard [3], but it has been intensively studied because of its great practical relevance [4], [5] and others. In recent years, the interest in the heuristic and hybrid approaches towards solving this problem has increased. These approaches give better results than the approaches based on constructive heuristics [6], [7] and [8].

There are different approaches that are used to solve the UCTP, for instance: constructive heuristics, meta-heuristics and constraints-based approaches. They are discussed in detail in the scientific literature [4], [9], [10], [11] and [12]. In addition to these approaches others are well known as well, for instance: multicriteria approaches, case-based reasoning, knowledge-based approaches and hyper-heuristic approaches [13].

A. Constraint-based approaches

In addition to the use of constraints in the constraint-based approaches, other supporting methods are used, such as: "Depth First Search", object-oriented modeling of graphs and trees, "backtracking", combined methods and genetic algorithms [14]. The experimental results show that it is possible for certain acceptable time to find good solutions that are close to the optimal one, but it refers only to timetables with a small number of events. This can be done by not considering temporary solutions that are not promising.

B. Graph-based approaches

Graph-based approaches show how the UCTP can be represented by a graph [4]. The graph coloring problem and its relationship with the UCTP are widely discussed in the scientific literature, for instance in [15].

C. Meta-heuristic and hyper-heuristic approaches

Meta-heuristic and hyper-heuristic approaches are methods of high level which are used to find the solution to problems with a large computational complexity. For instance, such are: "tabu search" [16]; "simulated annealing" [17]; "variable neighborhood search" [18] and "ant colony optimization" [19].

The purpose of these approaches is maximum satisfaction of the soft constraints. They are one of the most effective strategies for the practical solution to optimization problems. The published results indicate that the proposed methods find good solutions when they are used for UCTP. Their disadvantage is the need to set up additional parameters that control the performance of the algorithms.

D. Case-based reasoning and knowledge-based approaches

Case-based reasoning approaches (CBR) are characterized by the fact that additional heuristic methods are used. For instance, graphs in which the attributes of the vertices and the edges store more information about the interconnection between events. In this way, the algorithm that generates a timetable shall decide how to continue the process from here (or to improve the final solution) [12] and [13]. Knowledge-based approaches use an expert system of rules with pre-defined strategies (for instance, "Depth-first search" [20].

E. Population-based approaches

In solving UCTP quite often population-based approaches are used. The most commonly used algorithms of this type are genetic and memetic [21] and their modifications presented in [22]. The published results indicate that these approaches generate good acceptable solutions for a short time.

An analytical description of the real UCTP is presented in [23]. The proposed model includes parameters, vectors and matrices, which are used in solving the problem, as well as a function to evaluate the found solutions. The soft constraints are described by weights which provides greater flexibility in their analysis. The implementation of a genetic algorithm (GA) and a memetic algorithm (MA), as well as their computational complexity (respectively, quadratic for GA and cubic for MA), are presented in [24]. These algorithms are used to solve the real UCTP. The solutions found are evaluated according to the model presented in [23]. It is shown experimentally that for the same input data GA generates good solutions comparable to those obtained by solving the problem of the user – expert. Unlike GA, MA generates better solutions (for all test input data sets) but runs slower because of its higher computing complexity [24].

In [25] an approach in which the events are grouped in groups of the same size is used. Then, the best solution to a given order of the events in the first group is looked for. Similarly, the best solutions in the order of events in the other groups are looked for. In this way the best solution for a given group cannot be worse than the last best solution found for the previous group. The results obtained for some input data are the best ones found so far. For the other tested input data sets, the algorithm found solutions commensurate with those found by MA [24]. However, not all possible groupings of events have been investigated (and only a small number of multiples of the number of events) which motivated the authors to focus on this subject of study in this article.

II. An Event Grouping Based Algorithm

An event grouping based algorithm (EGB) to UCTP will be presented. All possible groupings of events in commensurate in size groups will be generated. The algorithm will search for the best solution for each successive order of events in each of the groups. It is necessary to determine how the number of groups affects the quality of the solutions found. As mentioned above (and described in [24]), for large size of the input data (on the order of several thousand events), the performance of the MA will take more computing time (due to the fact that more solutions should be found) in comparison with the algorithm, EGB which also will use the evaluation model presented in [23]. This algorithm is integrated into the updated version of the information system for the automated university course timetabling presented in [26].

Let N is a set of n events, i.e. $E = \{e_1, e_2, ..., e_n\}$, $n \geq 4$ and G is a set of m different ways of grouping these events, i.e. $G = \{g_1, g_2, ..., g_m\}$ such that $2 \leq m \leq \lfloor n/2 \rfloor$, or in other words it is necessary to establish at least two groups, as in any group, there are at least two events. The union of all groups of events gives the set E, i.e. $g_1 \cup g_2 \cup ... \cup g_m = E$, or in other words every event is in exactly one group, i.e. $g_i \cap g_j = \varnothing$, for $\forall\ i \neq j$. The cardinality of any two groups should not differ with more than one event, i.e., it must be satisfied:

$$\left\| g_i \right| - \left| g_j \right\| = \begin{cases} 0, & if\left(n \bmod m\right) = 0 \\ 1, & otherwise \end{cases}, \quad i \neq j \tag{1}$$

To satisfy (1) it is necessary that the n mod m groups (i.e. the remainder of dividing the n and m) have exactly the $\lfloor n/m \rfloor+1$ events (i.e. the quotient of the division of n and m without remainder). Some other interesting techniques using grouping of resources (not necessarily the events) are found in the scientific literature, for example in [27] and [28].

Below an example with 11 events and their distribution in 2, 3, 4 and 5 groups is presented.

TABLE I. Distribution of 11 Events into 2, 3, 4 and 5 Groups

$m = 2; \lfloor n / m \rfloor = 5; (n \bmod m) = 1; \lfloor n / m \rfloor + 1 = 6$											
$\boldsymbol{e_n}$	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}	e_{11}
$\boldsymbol{g_2}$	$\lvert g_1 \rvert = 6$						$\lvert g_2 \rvert = 5$				

$m = 3; \lfloor n / m \rfloor = 3; (n \bmod m) = 2; \lfloor n / m \rfloor + 1 = 4$											
$\boldsymbol{e_n}$	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}	e_{11}
$\boldsymbol{g_3}$	$\lvert g_1 \rvert = 4$				$\lvert g_2 \rvert = 4$				$\lvert g_3 \rvert = 3$		

$m = 4; \lfloor n / m \rfloor = 2; (n \bmod m) = 3; \lfloor n / m \rfloor + 1 = 3$											
$\boldsymbol{e_n}$	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}	e_{11}
$\boldsymbol{g_4}$	$\lvert g_1 \rvert = 3$			$\lvert g_2 \rvert = 3$			$\lvert g_3 \rvert = 3$			$\lvert g_4 \rvert = 2$	

$m = 5; \lfloor n / m \rfloor = 2; (n \bmod m) = 1; \lfloor n / m \rfloor + 1 = 3$											
$\boldsymbol{e_n}$	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}	e_{11}
$\boldsymbol{g_5}$	$\lvert g_1 \rvert = 3$			$\lvert g_2 \rvert = 2$		$\lvert g_3 \rvert = 2$		$\lvert g_4 \rvert = 2$		$\lvert g_5 \rvert = 2$	

After conducting the experiments and analyzing the obtained results it was found that the best solutions are not always generated when events are distributed in regular groups.

An implementation of the EGB algorithm will be presented in the Object Pascal (Delphi) language.

```
procedure EventGrouping(n: integer);
var
  m, g, r: integer;
  tg, tr, tn, tm: integer;
  flag: boolean;
  i, j, count: integer;
  from_index, to_index, best_index: integer;
  first, tmp: integer;
  eval, best_eval: single;
  p: array of integer;
  e: array of integer; //an array of events
  groups: array of integer; //an array of groups
  col, row: integer;
begin
```


https://sites.google.com/site/ijcsis/
ISSN 1947-5500

```
  setlength(p, n); //memory allocation for p
  setlength(e, n); //memory allocation for e
  for m := 2 to (n div 2) do //for each group
  begin
    g := n div m; //events in group
    r := n mod m; //undistributed events
    groups := nil; //deallocate groups array
    setlength(groups, m + 1); //allocate memory
    tg := g; //number of events
    tr := r; //undistributed events
    flag := false; //a boolean variable
    count := 0;
    tm := 1; //the first group
    groups.cells[1, 1] := 1;
    for tn := 1 to n do
    begin
      e.cells[tn, 1] := tm;
      tg := tg - 1; //an event is fixed
      count := count + 1; //the same as inc(count)
      if ((tg = 0) and (tr > 0) and (not flag)) then
      begin
        tg := 1;
        tr := tr - 1; //the same as dec(tr)
        flag := true;
        continue; //continue to the next iteration
      end;
      if (tg = 0) then
      begin
        groups.cells[2, tm] := tn;
        groups.cells[3, tm] := count;
        tm := tm + 1; //the same as inc(tm)
        tg := g;
        count := 0;
        if (tr > 0) then flag := false;
        if (tm <= m) then
          groups.cells[1, tm] := tn + 1;
      end;
    end; //for tn := 1 to n do
    for tm := 1 to m do //for each group
    begin
      from_index := groups.cells[1, tm];
      to_index := groups.cells[2, tm];
      best_eval := maxint; //init best_eval
      best_index := 0; //init best_index
      for i := from_index to to_index do
      begin
        LocalSearch; //call LocalSearch method
        if (eval < best_eval) then
        begin
          best_eval := eval;
          best_index := i;
        end;
        //move events from from_index to to_index
        //to the left one position
        first := p[from_index];
        for j := from_index to to_index - 1 do
          p[j] := p[j + 1];
        p[to_index] := first;
      end; //for i := from_index to to_index do
      tmp := p[1];
      for j := 1 to (best_index - from_index) do
        p[j] := p[j + 1];
      p[j] := tmp;
    end; //for tm := 1 to m do
  end; //for m := 2 to (n div 2) do
end; //end EventGrouping method
```

For each grouping *m* the EGB algorithm rearranges all events *n*. After each rearrangement of the events (in a group) the local search method is called which finds the best solution in this order of events. As the complexity of the *LocalSearch* method is the quadratic [24], for the proposed algorithm it is found out that there is a computational complexity $O = m.n^3$. In the General case the complexity is cubic which also depends on the number of groupings $m = n/2 - 1$. Finding a way to reduce the number of groupings will reduce the execution time of the EGB algorithm.

III. Experimental Results

The object of the study is an updated version of the integrated information system to university course timetabling. Its development and use are described in [26]. In the updated version of the system and EGB algorithm, that was presented above, was added (Fig. 1).

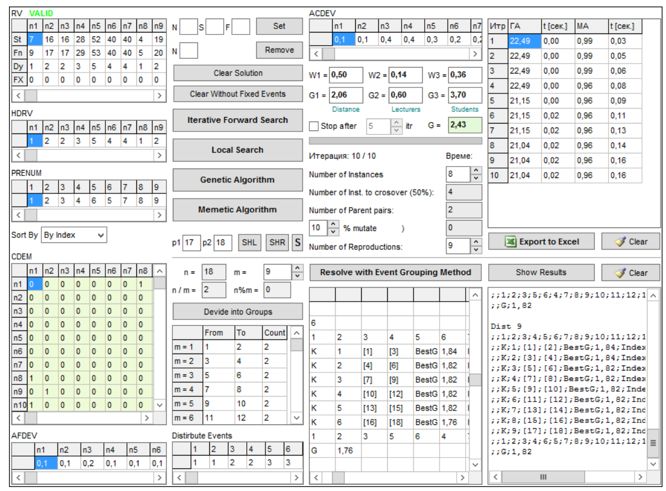


Figure 1. Working session with the updated version of the system.

With this system specific experiments to test the EGB algorithm with real data can be made.

The aim of the experiments was to determine the behavior of the algorithm on specific input data sets which are presented in [24]. For these input data sets there is already information concerning the algorithms used and the best solutions found. For some input data the EGB algorithm generated the best currently known solutions so far. In order to determine (experimentally) under what groupings of events the best results are received, all possible groupings will be generated.

A. Experimental Conditions

The experimental conditions for conducting the experiments are the following: PC with 64-bit Operating System Windows 10 Pro, x64-based processor and the following hardware configuration: Processor: Intel(R) Core(TM) i7-4712MQ CPU at 2.30 GHz; RAM memory: 8 GB DDR3 L.

B. Methodology of the experiment

To achieve the goals of the experiments three input data sets were used:

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

- Input data set DS_E90S175L29A18 with ninety events (90), one hundred and seventy-five students (175), twenty-nine lecturers (29) and eighteen auditoriums (18);
- Input data set DS_E130S274L37A22 with one hundred and thirty events (130), two hundred and seventy-four students (274), thirty-seven lecturers (37) and twenty-two auditoriums (22);
- Input data set DS_E273S549L62A39 with two hundred and seventy-three events (273), five hundred and forty-nine students (549), sixty-two lecturers (62) and thirty-nine auditoriums (39).

C. Experimental results

In Fig. II, the results of the EGB algorithm execution on input data set DS_E90S175L29A18 are shown. The events are sorted in order by index, weight, number and duration. This sequence was the same in all experiments.

TABLE II. RESULTS FOR DS_E90S175L29A18

m	Groups	Index	Weight	Number	Duration
2	2x45	9.758	7.422	8.545	**6.967**
3	3x30	9.312	**6.530**	7.453	8.002
4	2x23; 2x22	9.120	7.652	7.198	7.835
5	5x18	7.304	6.817	7.821	7.695
6	6x15	7.561	6.597	6.823	7.137
7	6x13; 1x12	7.618	7.469	8.207	7.487
8	2x12; 6x11	7.589	7.459	7.228	8.047
9	9x10	**7.018**	7.247	8.278	8.423
10	10x9	8.740	7.464	8.637	8.314
11	2x9; 9x8	9.365	7.491	8.604	8.418
12	6x8; 6x7	8.341	7.431	8.990	7.140
13	12x7; 1x6	7.598	7.502	**6.759**	7.264
14	6x7; 8x6	7.811	7.529	6.787	7.264
15	15x6	8.987	7.529	7.170	7.662
16	10x6; 6x5	9.107	7.518	7.170	7.922
17	5x6; 12x5	8.999	7.518	7.154	7.922
18	18x5	8.029	7.902	7.319	7.635
19	14x5; 5x4	8.085	7.902	7.319	7.647
20	10x5; 10x4	7.516	7.902	7.319	7.879
21	6x5; 15x4	8.879	7.902	7.319	7.879
22	2x5; 20x4	10.542	7.902	7.217	7.660
23	21x4; 2x3	9.936	7.718	8.882	7.791
24	18x4; 6x3	9.936	7.799	8.882	8.353
25	15x4; 10x3	9.931	7.853	8.905	8.353
26	12x4; 14x3	9.931	8.060	8.905	8.359
27	9x4; 18x3	9.931	8.060	8.905	8.359
28	6x4; 22x3	10.946	8.060	8.748	8.359
29	3x4; 26x3	10.946	8.060	9.924	8.359
30	30x3	8.371	8.421	8.349	9.090
31	28x3; 3x2	8.371	8.421	8.349	9.090
32	26x3; 6x2	8.371	8.421	8.349	9.090
33	24x3; 9x2	8.376	8.421	8.349	9.090
34	22x3; 12x2	8.376	8.421	8.366	9.090
35	20x3; 15x2	8.376	8.421	8.366	9.090
36	18x3; 18x2	8.376	8.421	8.366	9.090
37	16x3; 21x2	8.376	8.421	8.366	9.172
38	14x3; 24x2	8.387	8.719	8.366	9.172
39	12x3; 27x2	8.387	8.719	8.366	9.172
40	10x3; 30x2	8.387	8.719	8.361	9.172
41	8x3; 33x2	10.308	8.719	8.361	9.183
42	6x3; 36x2	10.308	8.719	8.361	7.504
43	4x3; 39x2	10.308	8.719	9.421	7.504
44	2x3; 42x2	11.100	8.071	9.700	8.194
45	45x2	11.100	8.769	9.239	8.194

The influence of the group number on the solution value for an input data set DS_E90S175L29A18 is shown in Fig. 2.

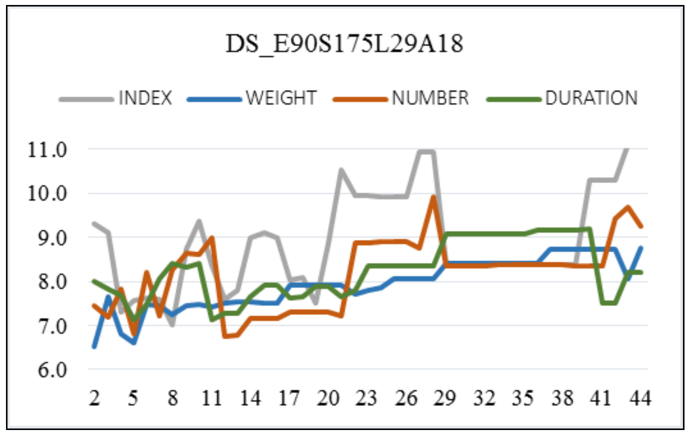


Figure 2. Influence of the group number (the x axis) on the solution value (the y axis) for DS_E90S175L29A18.

In Fig. III, the best results of the EGB algorithm execution on an input data set DS_E90S175L29A18 (for each sort criteria) are shown.

TABLE III. THE BEST RESULTS FOR DS_E90S175L29A18

By	Index	Weight	Number	Duration
Best	m=9: 7.018	m=3: 6.530	m=13: 6.759	m=2: 6.967

The influence of the sort criteria on the best solution value for an input data set DS_E90S175L29A18 is shown in Fig. 3.

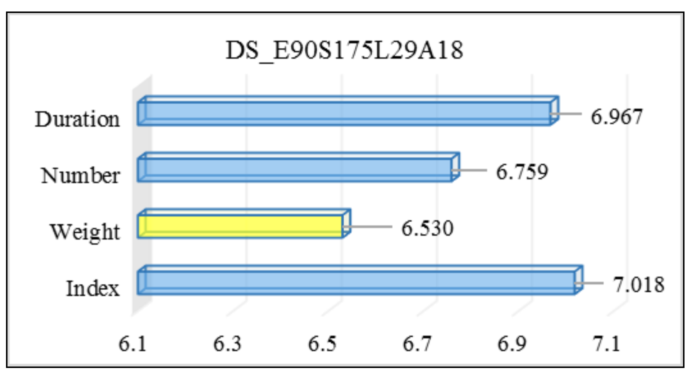


Figure 3. Influence of the sort criteria (the y axis) on the best solution value (the x axis) for DS_E90S175L29A18.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Fig. II, III, 2 and 3 show that for the input data set DS_E90S175L29A18 the best found solution is with a value of 6.530. The solution was obtained when the events were sorted by weight and divided into 3 groups (respectively with 30 events in each). Another good solution (with a value of 6.759) was found when the events were sorted by number and divided into 13 groups (12 groups with 7 events and a group with 6 events). When the events were sorted by index, the best found solution (with a value of 7.018) is the worst found solution of all other solutions found when sorting the events in the other three criteria.

In Fig. IV, the results of the EGB algorithm execution on input data set DS_E130S274L37A22 are shown.

TABLE IV. RESULTS FOR DS_E130S274L37A22

m	Groups	Index	Weight	Number	Duration
2	2x65	12.227	11.489	11.331	11.547
3	1x44; 2x43	12.762	11.177	11.200	11.147
4	2x33; 2x32	15.118	10.170	11.332	10.556
5	5x26	12.476	9.689	11.328	9.707
6	4x22; 2x21	11.824	10.283	11.659	10.820
7	4x19; 3x18	**10.070**	10.006	11.331	10.526
8	2x17; 6x16	11.692	**9.158**	10.552	10.882
9	4x15; 5x14	10.580	9.677	11.864	11.663
10	10x13	10.714	10.070	10.470	10.623
11	9x12; 2x11	10.878	9.787	10.703	12.108
12	10x11; 2x10	12.170	10.000	**10.239**	**8.958**
13	13x10	13.422	10.032	11.155	9.549
14	4x10; 10x9	13.254	10.093	11.376	9.658
15	10x9; 5x8	10.423	9.509	11.057	10.236
16	2x9; 14x8	11.306	9.628	12.298	10.685
17	11x8; 6x7	10.823	10.286	10.954	10.867
18	4x8; 14x7	11.683	10.297	11.491	11.592
19	16x7; 3x6	13.353	10.774	12.272	9.268
20	10x7; 10x6	13.797	10.774	12.299	9.307
21	4x7; 17x6	13.797	10.918	12.299	9.716
22	20x6; 2x5	12.620	11.130	11.339	10.035
23	15x6; 8x5	12.844	11.122	11.347	10.020
24	10x6; 14x5	12.214	10.842	11.347	10.020
25	5x6; 20x5	13.225	10.386	11.604	10.514
26	26x5	13.729	10.473	11.524	11.737
27	22x5; 5x4	13.729	10.481	11.535	11.737
28	18x5; 10x4	13.729	10.481	11.539	11.741
29	14x5; 15x4	13.800	10.893	11.539	11.752
30	10x5; 20x4	13.800	11.008	11.539	11.752
31	6x5; 25x4	15.548	11.148	11.651	11.878
32	2x5; 30x4	11.828	11.162	11.651	11.878
33	31x4; 2x3	12.894	11.031	13.265	12.413
34	28x4; 6x3	12.894	11.085	13.265	12.413
35	25x4; 10x3	12.894	11.085	12.920	12.413
36	22x4; 14x3	13.859	11.085	12.920	12.413
37	19x4; 18x3	13.859	11.608	12.920	12.317
38	16x4; 22x3	13.859	11.608	12.920	12.317
39	13x4; 26x3	13.859	11.591	12.920	12.317
40	10x4; 30x3	13.842	11.591	12.920	12.317
41	7x4; 34x3	13.842	11.591	12.812	12.294
42	4x4; 38x3	13.765	11.591	12.701	12.399
43	1x4; 42x3	13.765	11.591	13.041	11.401
44	42x3; 2x2	15.914	11.130	12.734	12.159
45	40x3; 5x2	16.032	11.130	12.734	12.159
46	38x3; 8x2	16.391	11.130	12.734	12.159
47	36x3; 11x2	16.303	11.130	12.734	12.159
48	34x3; 14x2	16.303	11.130	12.734	12.159
49	32x3; 17x2	16.288	11.130	12.734	12.159
50	30x3; 20x2	16.738	11.243	12.734	12.159
51	28x3; 23x2	16.738	11.162	12.734	12.51
52	26x3; 26x2	16.738	11.162	12.734	12.518
53	24x3; 29x2	16.738	11.162	12.734	12.518
54	22x3; 32x2	16.738	11.162	12.734	12.518
55	20x3; 35x2	16.738	11.162	12.734	12.518
56	18x3; 38x2	16.738	11.162	12.734	12.751
57	16x3; 41x2	16.738	11.162	12.734	12.751
58	14x3; 44x2	15.690	11.162	13.478	12.751
59	12x3; 47x2	15.690	11.162	13.478	12.751
60	10x3; 50x2	15.690	11.162	13.478	12.751
61	8x3; 53x2	15.690	11.162	13.478	12.751
62	6x3; 56x2	15.690	11.162	13.250	12.751
63	4x3; 59x2	15.690	10.457	13.250	12.751
64	2x3; 62x2	15.690	11.861	13.250	12.751
65	65x2	15.690	11.861	13.467	11.000

The influence of the group number on the solution value for an input data set DS_E130S274L37A22 is shown in Fig. 4.

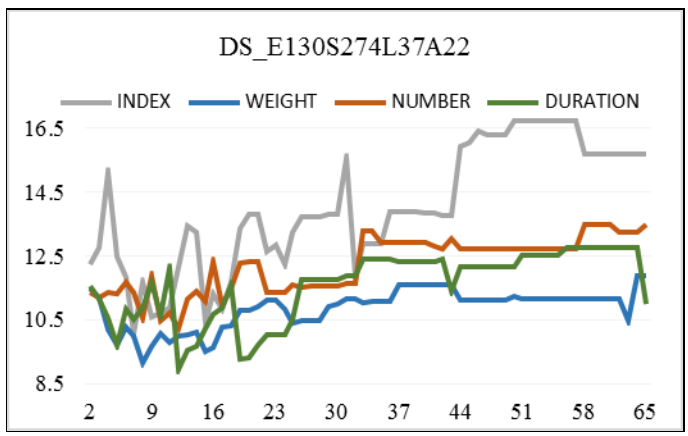


Figure 4. Influence of the group number (the x axis) on the solution value (the y axis) for DS_E130S274L37A22.

In Fig. V, the best results of the EGB algorithm execution on an input data set DS_E130S274L37A22 (for each sort criteria) are shown.

TABLE V. THE BEST RESULTS FOR DS_E130S274L37A22

By	Index	Weight	Number	Duration
Best	m=7: 10.070	m=8: 9.158	m=12: 10.239	m=12: 8.958

The influence of the sort criteria on the best solution value for an input data set DS_E130S274L37A22 is shown in Fig. 5.

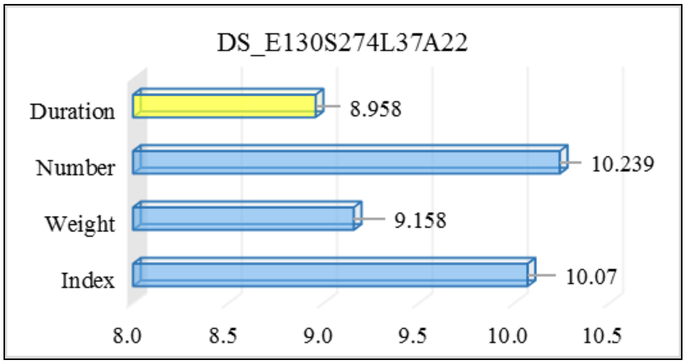


Figure 5. Influence of the sort criteria (the y axis) on the best solution value (the x axis) for DS_E130S274L37A22.

Fig. IV, V, 4 and 5 show that for the input data set DS_E130S274L37A22 the best found solution is with a value of 8.958. The solution was obtained when the events were sorted by duration and divided into 12 groups (10 groups with 11 events and 2 groups with 10 events). Another good solution (with a value of 9.158) was found when the events were sorted by weight and divided into 8 groups (2 groups with 17 events and 6 groups with 16 events). When the events were sorted by number, the best found solution (with a value of 10.239) is the worst found solution of all other solutions found when sorting the events in the other three criteria.

In Fig. VI, the results of the EGB algorithm execution on input data set DS_E273S549L62A39 are shown.

TABLE VI. RESULTS FOR DS_E273S549L62A39

m	Groups	Index	Weight	Number	Duration
2	1x137; 1x136	37.582	26.480	26.072	25.406
3	3x91	29.974	26.323	25.494	24.452
4	1x69; 3x68	34.971	24.072	23.133	23.163
5	3x55; 2x54	34.735	23.413	25.024	22.942
6	3x46; 3x45	31.980	**22.068**	25.073	21.861
7	7x39	30.382	23.387	23.183	21.745
8	1x35; 7x34	28.247	23.038	25.383	21.655
9	3x31; 6x30	30.747	23.781	23.771	22.522
10	3x28; 7x27	31.623	23.125	23.608	22.341
11	9x25; 2x24	27.140	22.632	24.100	23.104
12	9x23; 3x22	31.971	26.780	25.074	21.958
13	13x21	34.048	24.518	25.580	**20.978**
14	7x20; 7x19	27.902	23.419	23.987	21.707
15	3x19; 12x18	28.639	23.095	25.473	21.672
16	1x18; 15x17	31.885	23.891	24.587	22.413
17	1x17; 16x16	30.846	24.603	**22.948**	22.800
18	3x16; 15x15	30.847	24.853	23.099	22.195
19	7x15; 12x14	29.142	25.560	25.702	22.541
20	13x14; 7x13	**26.562**	24.867	25.303	22.536
21	21x13	29.473	24.310	23.600	22.953
22	9x13; 13x12	31.004	23.785	24.406	22.912
23	20x12; 3x11	33.172	25.287	23.727	24.327
24	9x12; 15x11	29.104	25.891	24.332	24.325
25	23x11; 2x10	32.128	25.226	25.073	23.270
26	13x11; 13x10	29.906	25.451	25.073	23.185
27	3x11; 24x10	27.928	25.506	24.942	23.032
28	21x10; 7x9	30.007	23.698	24.859	23.595
29	12x10; 17x9	32.714	23.802	23.758	23.871
30	3x10; 27x9	31.514	23.410	24.629	24.150
31	25x9; 6x8	35.509	24.879	25.688	23.917
32	17x9; 15x8	30.103	25.089	25.794	23.738
33	9x9; 24x8	32.117	24.983	25.867	24.795
34	1x9; 33x8	32.687	25.340	25.147	24.009
35	28x8; 7x7	33.098	23.990	25.698	25.021
36	21x8; 15x7	33.433	23.906	25.933	24.727
37	14x8; 23x7	36.308	24.069	26.082	25.001
38	7x8; 31x7	36.272	24.069	27.290	24.310
39	39x7	33.044	24.122	25.427	24.368
40	33x7; 7x6	33.046	24.120	25.404	24.368
41	27x7; 14x6	33,037	23,989	25,404	24,988
42	21x7; 21x6	32,646	23,766	25,404	24,988
43	15x7; 28x6	32,748	23,669	25,287	24,988
44	9x7; 35x6	33,324	23,721	25,284	25,585
45	3x7; 42x6	32,406	24,189	26,052	25,836
46 … 135	43x6; 2x5 … 3x3; 132x2	greater than 26.562	greater than 22.068	greater than 22.948	greater than 20.978
136	1x3; 135x2	35.686	26.363	29.085	29.026

The influence of the group number on the solution value for an input data set DS_E273S549L62A39 is shown in Fig. 6.

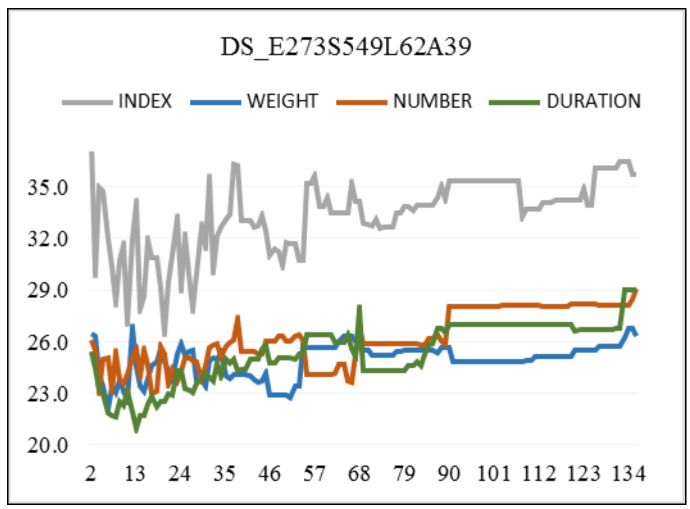


Figure 6. Influence of the group number (the x axis) on the solution value (the y axis) for DS_E273S549L62A39.

In Fig. VII, the best results of the EGB algorithm execution on an input data set DS_E273S549L62A39 (for each sort criteria) are shown.

TABLE VII. THE BEST RESULTS FOR DS_E273S549L62A39

By	Index	Weight	Number	Duration
Best	m=20: 26.562	m=6: 22.068	m=17: 22.948	m=13: 20.978

The influence of the sort criteria on the best solution value for an input data set DS_E273S549L62A39 is shown in Fig. 7.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

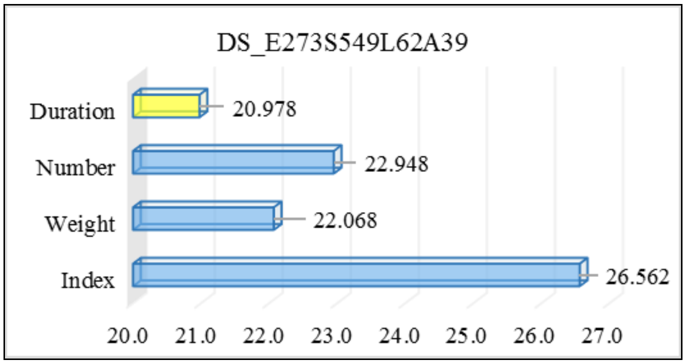


Figure 7. Influence of the sort criteria (the y axis) on the best solution value (the x axis) for DS_E273S549L62A39.

Fig. VI, VII, 6 and 7 show that for the input data set DS_E273S549L62A39 the best found solution is with a value of 20.978. The solution was obtained when the events were sorted by duration and divided into 13 groups (respectively with 21 events in each). Another good solution (with a value of 22.068) was found when the events were sorted by weight and divided into 6 groups (3 groups with 46 events and 3 groups with 45 events). When the events were sorted by index, the best found solution (with a value of 26.562) is the worst found solution of all other solutions found when sorting the events in the other three criteria.

IV. CONCLUSIONS

The best results, the sort criteria and the number of groups after five starts of EGB algorithm (for all input data sets) are shown in Fig. VIII.

TABLE VIII. THE BEST FIVE RESULTS FOR ALL INPUT DATA SETS

Input Data Set	**Start 1**	**Start 2**	**Start 3**	**Start 4**	**Start 5**
DS_E90 S175L29 A18	Weight $m = 3$ 6.530	Weight $m = 6$ 6.597	Number $m = 13$ 6.759	Number $m = 14$ 6.787	Weight $m = 5$ 6.817
DS_E130 S274L37 A22	Duration $m = 12$ 8.958	Weight $m = 8$ 9.158	Duration $m = 19$ 9.268	Duration $m = 20$ 9.307	Weight $m = 15$ 9.509
DS_E273 S549L62 A39	Duration $m = 13$ 20.978	Duration $m = 8$ 21.655	Duration $m = 15$ 21.672	Duration $m = 14$ 21.707	Duration $m = 7$ 21.745

The ratio between the best solutions and the sort criteria (according to number, weight and duration) is shown in Fig. 8.

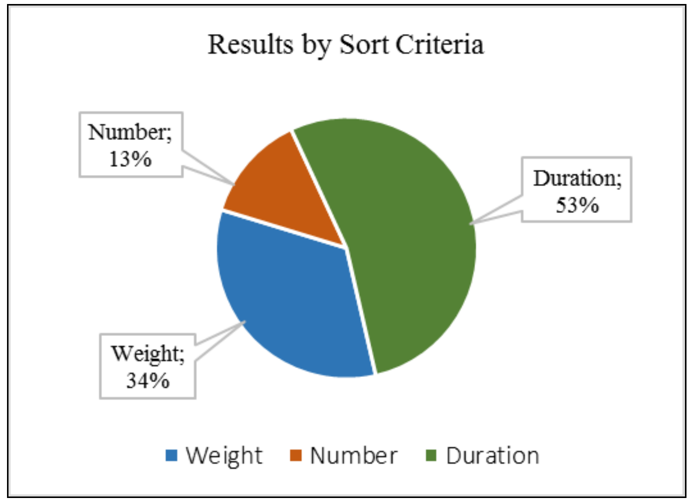


Figure 8. Ratio between the best solutions and the sort criteria.

Fig. VIII and 8 show that the EGB algorithm found 8 out of 15 best solutions (53%), when the events were sorted by duration. The other 5 solutions (34%) were obtained when the events were sorted by weight. And only 2 solutions (13%) were obtained when the events were sorted by number.

The ranges that contain the groups with the best results for all input data sets are shown in Fig. IX.

TABLE IX. RANGES OF THE GROUPS WITH THE BEST RESULTS

Input Data Set	*m*	**Range**	**Range calculated by m**
DS_E90S175L29A18	45	[3, ..., 14]	[m / 15.0, ..., m / 3.2]
DS_E130S274L37A22	65	[8, ..., 20]	[m / 8.1, ..., m / 3.25]
DS_E273S549L62A39	136	[7, ..., 15]	[m / 19.4, ..., m / 9.1]

The results obtained show that the range containing all the groups with the best solutions found is [m / 33.3, ..., m / 6.67] (summarized from the results for all input data sets).

After the analysis of the results the following conclusions can be made: 1) the EGB algorithm can be used to solve real UCTP; 2) the number of groups influences on the quality of the solutions found; 3) the number of the tested groups of events can be reduced considering only those that are within the range [m / 33.3, ..., m / 6.67].

The study presented in this paper may be extended in two guidelines: 1) optimization of the EGB algorithm from the point of view of computational complexity and 2) defining more precisely the range of tested groups through conducting additional experiments.

REFERENCES

[1] C. Gotlieb, "The construction of class teacher timetables," In C. M. Popplewell, Editor, IFIP Congress 62, pp. 73–77, North-Holland, 1963.

[2] A. Wren, "Scheduling, timetabling and rostering – a special relationship," The practice and theory of automated timetabling I: Selected papers from 1st International Conference (Patat I), Edinburgh, UK, Lecture notes in computer science 1153, Springer-Verlag, pp. 46-75, 1996.

[3] S. Even, A. Itai, and A. Shamir, "On the complexity of timetable and multi-commodity flow problems," SLAM Journal of computing, vol. 5, no. 4, pp. 691-703, 1976.

[4] D. Werra, "An introduction to timetabling," European journal of operational research, vol. 19, pp. 151–162, 1985.

[5] A. R. Komijan, and M. N. Koupaei, "A mathematical model for university course scheduling: a case study," International Journal of Technical Research and Applications, vol. 19, pp. 20-25, 2015.

[6] R. Chen, and H. Shih, "Solving university course timetabling problems using constriction particle swarm optimization with local search," Algorithms, vol. 6, pp. 227-244, 2013.

[7] M. S. Kohshori, and M. S. Abadeh, "Hybrid genetic algorithms for university course timetabling," International Journal of Computer Science Issues, vol. 9(2), pp. 446-455, 2012.

[8] E. K. Burke, J. Marecek, A. J. Parkes, and H. Rudova, "Decomposition, reformulation, and diving in university course timetabling," Computers & OperationsResearch, vol. 37, pp. 582-597, 2010.

[9] M. Carter, and G. Laporte, "Recent Developments in Practical Course Timetabling," The practice and theory of automated timetabling II: Selected papers from 2nd International conference (PATAT II), Toronto, Canada, Lecture notes in computer science 1408, Springer-Verlag, pp. 3-19, 1998.

[10] E. Burke, K. Jackson, J. Kingston, and R. Weare, "Automated University Timetabling: The State of the Art," The computer journal, 40 (9), pp. 565-571, 1997.

[11] A. Schaerf, "A survey of automated timetabling," Artificial intelligence review, vol. 13 (2), pp. 87-127, 1999.

[12] E. Burke, and S. Petrovic, "Recent research directions in automated timetabling," European journal of operational research, vol. 140 (2), pp. 266-280, 2002.

[13] S. Petrovic, and E. Burke, "University timetabling. handbook of scheduling: algorithms, models and performance analysis," chapter 45, (editor: J. Leung), CRC Press, 2004.

[14] L. Kang, and G. White "A logic approach to the resolution of constraints in timetabling," European journal of operational research, vol. 61, pp. 306-317, 1992.

[15] D. Werra, "Extensions of colouring models for scheduling purposes," European journal of operational research, vol. 92, pp. 474-492, 1996.

[16] D. Costa, "A tabu search for computing an operational timetable", European journal of operational research," vol. 76, pp. 98-110, 1994.

[17] M. Elmohamed, P. Coddington, and G. Fox, "A comparison of annealing techniques for academic course scheduling," The practice and theory of automated timetabling II: Selected papers from 2nd International conference (PATAT II), Toronto, Canada, Lecture notes in computer science 1408, Springer-Verlag, pp. 92-112, 1998.

[18] S. Abdullah, E. Burke, and B. McCollum, "An investigation of a variable neighborhood search approach for course timetabling," The proceedings of the 2nd Multidisciplinary international conference on scheduling: Theory and applications (MISTA 2005), New York, USA, p. 413-427, 2005.

[19] K. Patrick, and Z. Godswill, "Greedy ants colony optimization strategy for solving the curriculum based university course timetabling problem," British Journal of Mathematics & Computer Science, vol. 14(2), pp. 1-10, 2016.

[20] F. Partovi, and B. Arinze, "A knowledge-based approach to the faculty course assignment problem," Socio-economics planning science, vol. 29 (3), pp. 245-256, 1995.

[21] S. E. Soliman, and A. E. Keshk, "Memetic algorithm for solving university course timetabling problem," International Journal of Mechanical Engineering and Information Technology, vol. 3(8), pp. 1476-1486, 2015.

[22] R. Lewis, and B. Paechter, "New Crossover operators for timetabling with evolutionary algorithms," The 5th International conference on recent advances in soft computing (RASC 2004), Nottingham, UK, vol. 5, pp. 189-195, 2004.

[23] V. Kralev, "A model for the university course timetabling problem," International journal "Information technologies & knowledge", vol. 3 (3), pp. 276-289, 2009.

[24] V. Kralev, "A genetic and memetic algorithm for solving the university course timetabling problem," International journal "Information theories & applications", vol. 16 (3), pp. 291-299, 2009.

[25] V. Kralev, and R. Kraleva, "Variable neighborhood search based algorithm for university course timetabling problem," proceedings of the Fifth international scientific conference, FMNS-2013, pp. 202-214, 2013.

[26] V. Kralev, R. Kraleva, and N. Siniagina, "An integrated system for university course timetabling," Proceedings of the third international scientific conference – FMNS2009, vol. 1, pp. 99-105, 2009.

[27] R. P. Badoni, D. K. Gupta, and P. Mishra, "A new hybrid algorithm for university course timetabling problem using events based on groupings of students," Computers & industrial engineering, vol. 78, pp. 12–25, 2014.

[28] E. Falkenauer, "Genetic algorithms for grouping problems," New York: Wiley, 1998.

Digital Image Watermarking Using DCT And DWT To Improve Robustness

Priyanka Rani, Anupam Singh
Computer Science and Engineering
Shri Ramswaroop Memorial University
Lucknow, India

Avinash Kumar Singh
Computer Science and Engineering
Shri Ramswaroop Memorial University
Lucknow, India

Abstract— **Watermarking is the concept that provides protection in digital multimedia. This paper uses Discrete Wavelet Transform (DWT), Singular Value Decomposition (SVD) and Discrete Cosine Transform (DCT) concept for watermarking and extraction purpose. In result analysis we analyze extracted image from watermarked image after applying different attacks (like rotation, Gaussian noise, average filter attack, low pass filter, high pass filter, salt and pepper, Histogram Equalization etc). We find that this concept is robust against these types of attacks and provide high security.**

Keywords- Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Singular Value Decomposition (SVD), Cover Image, Watermark Message.

I. Introduction

In Today's digital world security has become an important issue both for owner and service provider [2], Watermarking is recognized as a major technology that has been developed to protect digital data (primarily images, audios and videos) from illegal manipulations [3]. Watermarking is the process of altering the cover work by embedding a watermark message into the cover work, watermarks are inconspicuous, and goes through the same transformation as the cover work. Later this watermark can be extracted from the cover work for security purpose. Watermark embedding can broadly perform into two domains [9]:

Spatial domain: Watermark embedding is achieved by altering the pixel values of the cover image.

Frequency domain: Watermark embedding is achieved by altering the transformed coefficients (by apply any transformation like DFT (Discrete Fourier Transformation), DCT (Discrete Cosine Transformation) etc.) of the cover image.

Watermark extraction or detection algorithms are classified on the basis of information needed by the detector [7]:

Non-blind scheme: Both Cover image and watermark message is needed.

Semi-blind scheme: Watermark message and watermark bit sequence are needed.

Blind scheme: Only watermark message is needed.

In Spatial domain we achieved watermarking [8] by embedding the watermark message at the least significant bits (LSB) of the cover image because LSB bits contains less information about the cover image , So it doesn't so much degrade the visual quality of the cover image. But this type of watermarking doesn't show robustness (ability to detect the watermark after common signal processing operations.) against attacks like cropping, scaling, compression etc.

To achieve better watermarking, which can show robustness against different types of attacks we turn towards transformed domain (frequency domain) [4-6]. In this paper we achieve watermarking using Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD).

Discrete Cosine Transform (DCT): DCT transforms a signal or image from the spatial domain to the frequency domain. Two dimensional discrete cosine transform (2D-DCT) is defined as:

$$f(u,v) = w(u)w(v)\sum_{x=0}^{N-1} C(x,y)\cos\left(\frac{\pi(2x+1)u}{2N}\right)\cos\left(\frac{\pi(2y+1)v}{2N}\right) \quad (1)$$

And corresponding 2D IDCT is defined as:

$$C(x,y) = \sum_{x=0}^{N-1} w(u)w(v)f(u,v)\cos\left(\frac{\pi(2x+1)u}{2N}\right)\cos\left(\frac{\pi(2y+1)v}{2N}\right) \quad (2)$$

Where u,v=0, 1, 2,…N-1.

Singular Value Decomposition (SVD): Every real matrix say A can be decomposed into a three matrices $A = U1*S1*V1^T$, where U1 and V1 are orthogonal matrices, $U1^TU1 = I$, $V1^TV1 = I$, and S1 = diagonal (λ1, λ2,

...). The diagonal S1 are known by singular values of A, columns of U1 are left singular vectors of A, and the columns of V1 are right singular vectors of A. This decomposition is known as the Singular Value Decomposition (SVD) of A [1].

Discrete Wavelet Transform (DWT): Discrete Wavelet Transform allows analysis in both frequency and spatial domain. For the given input function *f(n)* where, *n*=0,1,2,....,*M*-1 Forward DWT uses two basis functions(Transformation kernel).

Scaling Function Term:

$$W_\varphi(j_0,k) = \frac{1}{\sqrt{M}} \sum_n f(n)\ \varphi_{j0,k}(n) \qquad (3)$$

Where

$$\varphi_{i,k} = 2^{j/2}\varphi(2^j n - k)$$

DWT scaling term behave as low pass filters. It gives us approximation coefficients.

Wavelet Function Term:

$$W_\psi(j,k) = \frac{1}{\sqrt{M}} \sum_n f(n)\Psi j,k(n) \quad \text{for } j \geq j0 \qquad (4)$$

Where

$$\psi_{j,k} = 2^{j/2} \quad \psi(2^j n - k)$$

DWT wavelet vectors behave as high pass filters. And it gives us Detailed coefficients.

Corresponding Inverse Discrete Wavelet Transform (IDWT) can be written as:

$$f(n) = \frac{1}{\sqrt{M}} \sum_k W_\varphi(j_0,k)\ \varphi_{j0,k}(n) + \sum_{j=0}^{\infty} \sum_k W_\varphi(j.k)\ \psi_{j,k}(n) \qquad (5)$$

We can compute the 2-D wavelet transform by simply computing 1-D FDWT along rows and then along the resulting columns. Filters used in FDWT are known as analysis filter bank and filter used in IDWT are known as synthesis filter bank. After applying DWT transform our image will divided into four frequency sub-bands (LL LH HL HH) [2].

In spatial domain method watermark [8] is embedded directly by modifying the pixel values of cover image, but problem here is that if we cut a portion or compress watermarked image then watermark will not be extracted due to the loss of bits.

In most of watermarking techniques [4-6], watermark embedded into the frequency domain instead of the spatial domain for the robustness of the watermarking mechanism. The left upper corner 8X8 DCT coefficients of cover image are modified in zig-zag order to hide the watermark by using the embedding rule, this limits the frequency we can't get more than 16 level frequencies.

Only spatial correlation of the pixels inside the single 2-D block is considered and the correlation from the pixels of the neighboring blocks is neglected Impossible to completely de-correlate the blocks at their boundaries using DCT [8].

As technology advances more security demands, so DWT based watermarking comes in [10], through this we transfer more confidential data through internet. Here we decompose image into 4 sub-bands, we can further decompose these sub-bands according to our application need, and this gives higher security as our data is present in small area of image. No need to divide the input coding into non-overlapping 2-D blocks, it has higher compression ratios avoid blocking artifacts. This allows good localization both in time and spatial frequency domain. In DWT transformation of the whole image introduces inherent scaling.

Mei Jiansheng [3] propose combination of DCT and DWT here DCT of watermark image is find then based on mathematical rule these are fed into high frequencies of DWT of cover image, but problem here is not good quality watermark on extraction since large changes occur in values of DCT while extraction and when we apply low pass filter then we can't get watermark back.

DWT-SVD [2] based watermarking show good result.

In this paper for watermarking we have used DCT, DWT, SVD so that our result is robust various attacks and have high security.

II. PROPOSED METHODOLOGY

In this paper for Watermarking we applied the 2D DWT transform on cover image and decompose it into 2-level of DWT. HH sub-band of 2nd level of LL and LL sub-band of 2nd level of HH is converted into corresponding singular values say S1 and S2 these SVD values are modified with SVD of watermark message.

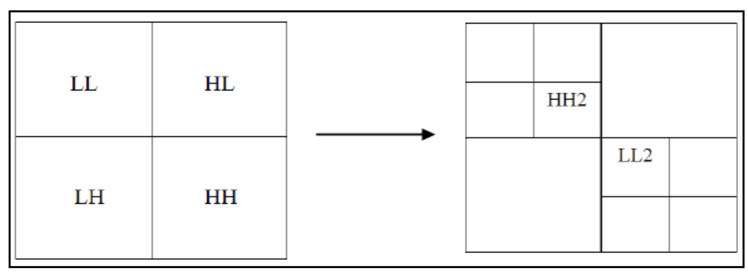


Figure 1: 2nd level 2-D DWT decomposition

Algorithm for Embedding:

1. Read the cover image, convert it into gray scale image.
2. Apply 2D DWT on cover image for 2 levels.
3. Apply SVD ($A = U1*S1*V1^{T}$) on HH sub-band of 2nd level of LL and LL sub-band of 2nd level of HH and calculate Singular values say S_i.
4. Read watermark image, convert it into gray scale image.
5. Apply DCT on watermark message image and then find its SVD say Ss_i.
6. On the HH2 and LL2 part as shown in figure 1, apply the Embedding concept:

   $S_i' = S_i + \eta * Ss_i$, where i=1, 2….n
7. Obtain the modified DWT coefficients:

   $A' = U1 * S_i' * V1^T$

8. Apply inverse DWT for 2 levels using the modified DWT coefficients to produce the watermarked cover image.

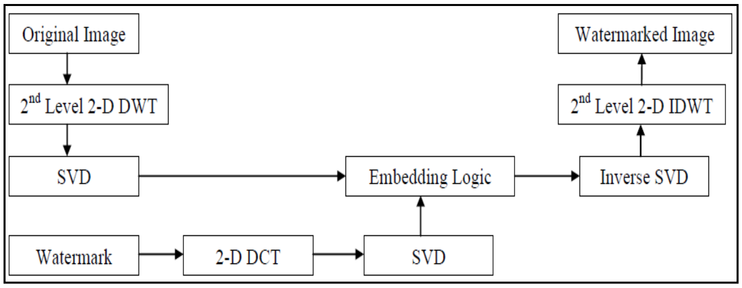


Figure 2: Embedding process of watermark

Algorithm for Extraction:

1. Read the embedded cover image, convert it into gray scale image.
2. Apply 2 level 2D DWT on embedded cover image.
3. Apply SVD ($A' = U1 * S_i' * V1^{T}$) on HH sub-band of 2nd level of LL and LL sub-band of 2nd level of HH say RS_i.
4. Read the cover image and apply 2- level 2D DWT on this image.
5. Apply SVD on HH sub-band of 2nd level of LL and LL sub-band of 2nd level of HH.
6. On the HH2 and LL2 part of embedded cover image and cover image, apply the Extracting concept:

   $RSs_i = (RS_i - S_i')/\eta$, where i=1, 2…..n.
7. Apply inverse SVD using extracted Singular component like $U1 * RS_{s*} V1^{T}$ and construct the extracted watermark image.

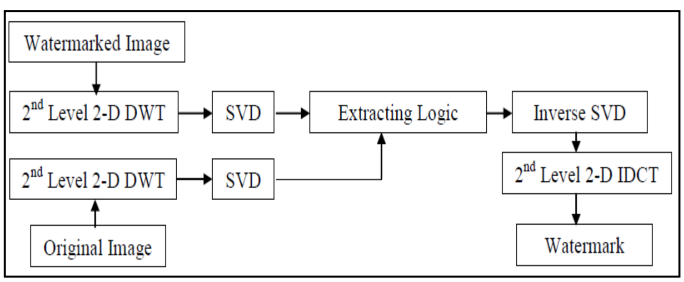


Figure 3: Extraction process of watermark

III. Experiment And Simulation

Table 1 shows the 300X300 gray scale cover image of 'Dog',the 75X75 gray scale watermark of 'Albert Einstein' and 300X300 watermarked image of 'Dog'. In this experiment scaling factor η is 0.01.

A. Result Analysis

The DWT, DCT, SVD based watermarking scheme tested using 14 attacks shown in table 2. The DWT is performed using the Haar wavelet filter. The chosen attacks are Average filter, Cropping, Gaussian Filter, Gaussian noise, High Pass Filter, Negative Laplacian, Poisson Attack, Positive Laplacian, Resizing, Rotation, Salt and Pepper noise, Weighted Average, Histogram Equalization.

Table 1: Cover image, Watermark Image and Watermarked Image

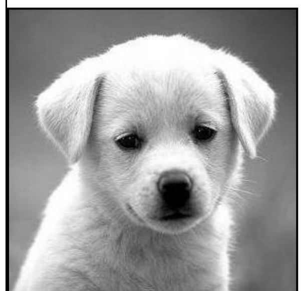
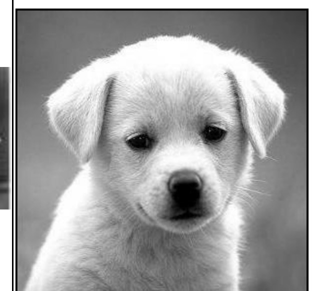
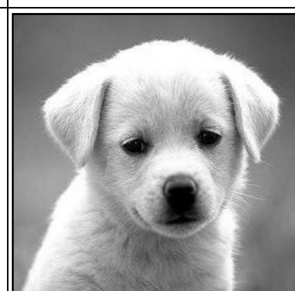
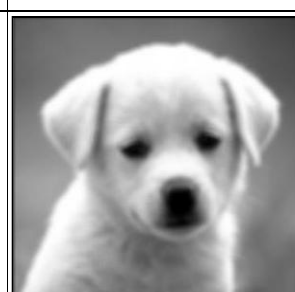
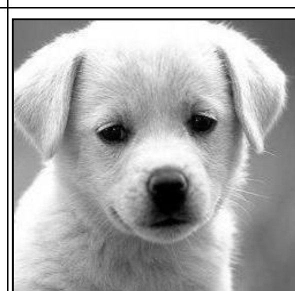
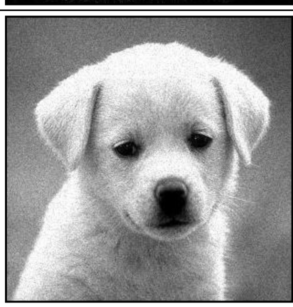
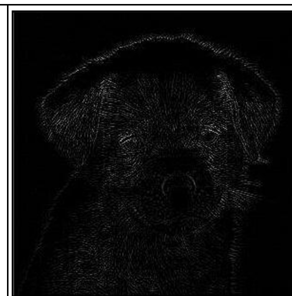
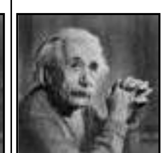
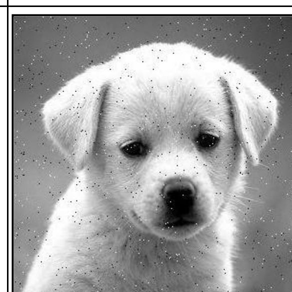
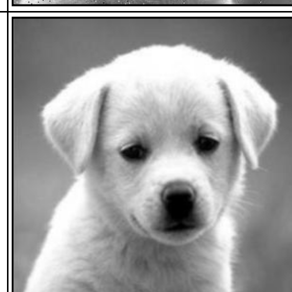
Cover Image 'Dog'	**Watermark 'Albert Einstein'**	**Watermarked Image**
		

Table 2: Watermark constructed in presence of noise from HH2 of LL and LL2 of HH look different for each attack.

Attacks	**Embedded Image in presence of different attack**	**Extraction of watermark from HH2 of LL**	**Extraction of watermark from LL2 of HH**
No Attack			
Average Filter Attack			
Cropping			

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Gaussian Filter			
Gaussian Noise(mean =0,variance-0.03)			
High Pass Filter			
Negative Laplacian Attack			
Poisson Attack			

Positive Laplacian Attack			
Resize Attack			
Rotation Attack(90^0)			
Salt and Pepper Noise Attack(noise density=0.01)			
Weighted Average Filter			

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Histogram Equalization Attack			

Watermark constructed from 2 sub-bands look different for each attack. Extraction of watermark HH2 of LL showing better result in case of high pass filter, negative laplacian, positive laplacian, histogram equalization and Extraction of watermark LL2 of HH showing better result in case of no attack, average filter, cropping, Gaussian filter, Gaussian noise, Poisson attack, resizing, rotation, salt and pepper and weighted average.

IV. Conclusion

Watermarks embedded in the low frequencies (LL sub-band) are resistant to one group of attacks, and watermarks embedded in high frequencies (HH sub-band) are resistant to another group of attacks. The watermarked message is embedded in 2 sub-bands so it becomes difficult to remove the watermark from embedded image. Here in this experiment the result is not up to the mark against some attacks like cropping and resizing, in future work we can try to improve the result by using watermark information from those 2 sub-bands, in some way.

References

[1] Alexander Sverdlov, Scott Dexter and Ahmet M. Eskicioglu, Robust Dct-Svd Domain Image Watermarking For Copyright Protection: Embedding Data In All Frequencies

[2] Emir Ganic and Ahmet M. Eskicioglu, Robust DWT-SVD Domain Image Watermarking: Embedding Data in All Frequencies, MM&SEC'04, September 20–21, 2004, Magdeburg, Germany. Copyright 2004 ACM 1-58113-854-7/04/0009

[3] Mei Jiansheng, Li Sukang and Tan Xiaomei, A Digital Watermarking Algorithm Based On DCT and DWT, Proceedings of the 2009 International Symposium on Web Information Systems and Applications (WISA'09) Nanchang, P. R. China, May 22-24, 2009, pp. 104-107

[4] H. Harrak, T. D. Hien, Y. Nagata, and Z. Nakao, DCT watermarking optimization by genetic programming, Advances in Soft Computing, 2006, pp. 347-351.

[5] C. C. Chang, P. Y. Lin, and J. S. Yeh, Preserving robustness and removability for digital watermarks using subsampling and difference correlation, Information Sciences, 2009, pp. 2283-2293.

[6] C. M.. Kung, J. H. Jeng, and T. K. Troung, Watermark technique using frequency domain, Proc. of the 14th IEEE International Conference on Digital Signal Processing (DSP2002), July. 2002, pp. 729-731.

[7] A. Sverdlov, S. Dexter, and A. Eskicioglu, Robust DCT SVD domain image watermarking for copyright protection: Embedding data in all frequencies, International Multimedia Conference, Germany, pp. 166-174, 2004.

[8] Preeti Parashar and Rajeev Kumar Singh, A Survey: Digital Image Watermarking Techniques, International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 7, No. 6 (2014), pp. 111-124.

[9] Kudratpreet Kaur, Malkit Singh, A study on digital watermarking algorithms, International Journal of Engineering and Innovative Technology (IJEIT) Volume 4, Issue 11, May 2015

[10] Radhika v. Totla, K.S.Bapat, Comparative Analysis of Watermarking in Digital Images Using DCT and DWT, International Journal of Scientific and Research Publications, Volume 3, Issue 2, February 2013 1 ISSN 2250-3153

AUTHORS PROFILE

Priyanka Rani M.Tech Computer Science and Engineering student, from Shri Ramswaroop Memorial University, Lucknow-Deva Road, Uttar Pradesh, India. Research Interest in Image Security, Pattern Recognization, Digital Image Processing.

Anupam Singh M.Tech Computer Science and Engineering student, from Shri Ramswaroop Memorial University, Lucknow-Deva Road, Uttar Pradesh, India. Research Interest in Pattern Recognization, Digital Image Processing, Artificial Intelligence.

Avinash Kumar Singh, Assistant Professor in Computer Science and Engineering Department at Shri Ramswaroop Memorial University, Lucknow-Deva Road, Uttar pradesh, India. Research Interest in Pattern Recognization, Image Security, Artifical Intellengence.

IJCSIS REVIEWERS' LIST

CALL FOR PAPERS
International Journal of Computer Science and Information Security

IJCSIS 2016
ISSN: 1947-5500
http://sites.google.com/site/ijcsis/

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc,), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on
its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embeded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at http://sites.google.com/site/ijcsis/authors-notes .

ISSN 1947 5500
http://sites.google.com/site/ijcsis/